Advances for Biomarker Discovery in Neuroproteomics using Mass Spectrometry

From Method Development to Clinical Application

MARCUS O.D. SJÖDIN
Proteins offer a prominent group of compounds which may be ubiquitously affected in disease and used as biomarkers for early diagnosis, assessing treatment or drug development. Clinical proteomics aim to screen for protein biomarkers by a comprehensive analysis of all proteins expressed in a biological matrix during a certain pathology. Characterization of thousands of proteins in a complex biological matrix is from an analytical point of view a challenging task. Hence, sophisticated methods that are sensitive, specific and robust in a high-throughput manner are required. Mass spectrometry (MS) is able to perform this to a wide extent is.

A prominent source for finding protein biomarkers related to neurological diseases is the central nervous system (CNS) due to close proximity of the pathogenesis. Neuroproteomic analysis of CNS tissue samples is thus likely to reveal novel biomarkers. Cerebrospinal fluid (CSF) bathes the entire CNS and offers a good balance between clinical implementation and usefulness. Both matrices put further requirements on the methodology due to a high dynamic range, low protein concentration and limited sample amount.

The central objective of this thesis was to develop, assess and utilize analytical methods to be used in combination with MS to enable protein biomarker discovery in the CNS. The use of hexapeptide ligand libraries was exemplified on CSF from patients with traumatic brain injury and demonstrated the ability to compress the dynamic range to enable protein profiling in the order of mg/mL to pg/mL. Further, a method based on cloud-point extraction was developed for simultaneous enrichment and fractionation of hydrophobic/hydrophilic proteins in brain tissue. Comparison between label and label-free MS based strategies were carried out, mimicking the true conditions with a few differentially expressed proteins and a bulk of proteins occurring in unchanged ratio. Finally, a clinical application was carried out to explore the molecular mechanism underlying the analgesic effect of spinal cord stimulation (SCS) in patients with neuropathic pain. The CSF concentration of Lynx1 was found to increase upon SCS. Lynx1, acting as a specific modulator of the cholinergic system in the CNS, may act as a potential important molecular explanation of SCS-induced analgesia.

**Keywords:** Mass Spectrometry, Biomarker, Proteomics, Central Nervous System, Cerebrospinal Fluid, Traumatic Brain Injury, Cloud-Point Extraction, Neuroproteomics, Relative Quantification, Spinal Cord Stimulation, Neuropathic Pain

_Marcus O.D. Sjödin, Uppsala University, Department of Chemistry - BMC, Analytical Chemistry, Box 576, SE-751 23 Uppsala, Sweden._

© Marcus O.D. Sjödin 2012

ISSN 1651-6214
ISBN 978-91-554-8457-6
urn:nbn:se:uu:diva-180109 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-180109)
“You cannot understand what you cannot measure”

Peter Roepstorff

To my family
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
Author’s contribution

**Paper I:** Took part in planning and performed most of the experiments. Performed the data handling and wrote the paper.

**Paper II:** Performed parts of the MS experiments and participated in writing the paper.

**Paper III:** Performed parts of the MS experiments and participated in writing the paper.

**Paper IV:** Took part in planning the study. Performed most of the experiments, data handling and wrote the paper.

**Paper V:** Took part in planning the study. Performed most of the experiments, data handling and wrote most of the paper.
Papers not included in the thesis


Contents

1. Introduction ........................................................................................................... 13

2. Proteomics ........................................................................................................... 15
   2.1 Clinical proteomics ....................................................................................... 15
   2.2 Analytical strategies ..................................................................................... 17

3. Biological Matrices .............................................................................................. 19
   3.1 Cerebrospinal fluid ....................................................................................... 20
      3.1.1 Physiology ............................................................................................ 20
      3.1.2 Proteome .............................................................................................. 21
   3.2 Brain tissue .................................................................................................... 21

4. Sample preparation .............................................................................................. 23
   4.1 Hexapeptide ligand library fractionation ..................................................... 23
   4.2 Cloud-point extraction .................................................................................. 25
   4.3 Immunoaffinity fractionation ......................................................................... 28

5. Separation techniques ......................................................................................... 30
   5.1 Liquid chromatography .................................................................................. 30
      5.1.1 Reversed phase liquid chromatography .............................................. 31
      5.1.2 Miniaturization of liquid chromatography .......................................... 32
   5.2 Electrophoresis .............................................................................................. 33
      5.2.1 Isoelectric focusing ............................................................................... 33
      5.2.2 Gel electrophoresis ................................................................................ 34

6. Mass spectrometry .............................................................................................. 36
   6.1 Ionization techniques ..................................................................................... 36
      6.1.1 Electrospray ionization .......................................................................... 36
      6.1.2 Matrix assisted laser desorption ionization .......................................... 39
   6.2 Mass analyzers .............................................................................................. 40
      6.2.1 Quadrupole ............................................................................................. 41
      6.2.2 Quadrupole ion trap ............................................................................... 41
      6.2.3 Fourier transform ion cyclotron resonance .......................................... 42
      6.2.4 Time-of-flight ......................................................................................... 42
   6.3 Fragmentation and sequencing of peptides ................................................... 43
   6.4 Bioinformatics ............................................................................................... 44
   6.5 Protein Quantification by Mass Spectrometry .............................................. 45
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&amp;2DGE</td>
<td>1&amp;2 dimensional gel electrophoresis</td>
</tr>
<tr>
<td>AC</td>
<td>affinity chromatography</td>
</tr>
<tr>
<td>BBB</td>
<td>blood-brain barrier</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CPE</td>
<td>cloud-point extraction</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DML</td>
<td>dimethyl labeling</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FQR</td>
<td>false quantification rate</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>HLL</td>
<td>hexapeptide ligand library</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HSA</td>
<td>human serum albumin</td>
</tr>
<tr>
<td>i.d.</td>
<td>internal diameter</td>
</tr>
<tr>
<td>IDA</td>
<td>information dependant acquisition</td>
</tr>
<tr>
<td>IEF</td>
<td>isoelectric focusing</td>
</tr>
<tr>
<td>IPG</td>
<td>immobilized pH gradient</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>isobaric tag for relative and absolute quantification</td>
</tr>
<tr>
<td>LF</td>
<td>label-free</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption ionization</td>
</tr>
<tr>
<td>MRM</td>
<td>multiple reaction monitoring</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass/charge</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>Q</td>
<td>quadrupole</td>
</tr>
<tr>
<td>QIT</td>
<td>quadrupole ion trap</td>
</tr>
<tr>
<td>RP</td>
<td>reversed phase</td>
</tr>
<tr>
<td>SAS</td>
<td>subarachnoid space</td>
</tr>
<tr>
<td>SCS</td>
<td>spinal cord stimulation</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TBI</td>
<td>traumatic brain injury</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight</td>
</tr>
</tbody>
</table>
1. Introduction

Historically, a diagnosis of a disease was made solely on the basis of a physical trait or a physiological observation. Today, a diagnosis is often made in conjunction with a laboratory test of measuring an objective biomarker. A biomarker is by definition; “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”\(^1\). Molecular biomarkers can be of any type of the different classes of compounds circulating in the body e.g. lipids, carbohydrates or metabolites. Consequently, many different methodological approaches may be adopted for biomarker discovery. Proteins offer a prominent group of compounds which may be ubiquitously affected in disease and hold promise for biomarker discovery\(^2\) (Chapter 2). The analysis of proteins present many analytical challenges such as high dynamic range and high structural diversity often encountered in biological samples. The central nervous system (CNS) is an isolated entity consisting of the brain and spinal cord which act as a prominent repository of potential protein biomarkers for neurological diseases\(^3\) (Chapter 3). Protein biomarker discovery within the CNS present an additional challenge due to the invasive procedures for sample collection as well as low protein concentration (Chapter 3). Specific and sensitive methods are required to obtain relevant information from material originating from this unique matrix. Initially, sample preparation to enrich proteins from the biological matrix is needed and is the first important step in the analysis (Chapter 4). The complexity of the sample needs to be reduced before detection which may be accomplished by different separation methods such as chromatography or electrophoresis (Chapter 5). Finally, the detection is carried out by mass spectrometry (MS) which is currently the only method capable of meeting the demands of specificity and sensitivity required for high throughput protein biomarker discovery (Chapter 6).

This thesis is based on Paper I-V which considers advancements in protein biomarker discovery in the CNS by neuroproteomics. The use of MS has consistently been used in all papers. Paper I-III mainly deals with the assessments of two sample preparation techniques for protein enrichment. In Paper I, the use of hexapeptide ligand libraries (HLL) containing millions of affinity baits which theoretically have a binding capacity for every possible protein was used for extensive mapping of the CSF proteome. The technique was exemplified from patients with traumatic brain injury (TBI). In Paper
temperature-induced cloud-point extraction (CPE) was explored for the simultaneous extraction, fractionation and concentration of hydrophobic and hydrophilic proteins from brain tissue. A quantitative evaluation of the optimized CPE method was then performed in Paper III. However, an efficient sample preparation technique alone is not sufficient for a successful biomarker discovery as all steps in the analytical chain of events are of equal importance. In Paper IV, three common MS based methods using shotgun proteomics for relative protein quantification were objectively evaluated. Different strengths and weaknesses were found for each of the three methods and the method of choice should be selected accordingly. The optimal MS method found in Paper IV was then utilized in Paper V. Spinal cord stimulation (SCS) is an effective treatment for patients with chronic neuropathic pain. The neurochemical mechanisms by which SCS produces an analgesic effect however still remain elusive. An unbiased protein biomarker discovery experiment was performed in Paper V and did yield a panel of differentially regulated proteins that may be of importance for the mechanistic explanation of SCS. The focus of each paper is put into context of the general phases of clinical neuroproteomic research in Figure 1.

Figure 1. The goal of clinical neuroproteomic research is ultimately to improve patient care. This thesis is based on the early analytical phases of this goal. Paper I-III mostly considers sample preparation for protein enrichment. The focus of Paper IV is protein analysis by MS but this detection method has been used in all papers. This thesis concludes with a clinical application in Paper V.
2. Proteomics

Proteomics is the large-scale study of the proteome, the entire protein complement of the genome expressed in a cell, tissue or organism at a given time point\(^6\). Proteomics aim to characterize molecular and cellular dynamics in protein expression and function on a global (simultaneous) level or targeting a panel of proteins. Proteins are the main functional output from the genome but the genetic code alone is insufficient to indicate which proteins are to be expressed, in what quantity and in what form. The transcriptome, the entire complement of mRNA transcripts transcribed from the genome may give some insights into what proteins might be expressed\(^7\). But in order to fully characterize the proteome, this needs to be done on a protein level. This however, comes with a higher complexity because unlike the genome which may be considered to be rather static, the proteome is highly dynamic. The proteome varies between cell types and fluctuates in response to physiological stimuli, changes in the extracellular milieu and disease\(^7\). Further complexity of the proteome comes from alternative splicing and post-translational modification such as, phosphorylation or acetylation of which there are about 200 different types described\(^8\). Taking this into account and that the human genome consist of about 21,000 genes, the total number of all structurally and functionally different proteins may very well lie in the order of a million\(^9\). Additionally, there is currently no global method of amplifying proteins as there is in comparison to transcriptomics and genomics which puts further demands on the analytical methodology.

2.1 Clinical proteomics

The observation that the proteome changes due to a disease makes it especially interesting from a clinical perspective. The pathological changes that lead to disease onset and progression may reflect itself in the proteome, the study of protein expression profiles may provide insights about the pathogenesis and development of effective strategies for treatment and early diagnosis\(^10\). Clinical proteomics aim to identify and quantitatively find alterations of protein biomarkers by directly analyzing the protein content in body fluids and tissues\(^2\). Proteins with altered concentration may serve as diagnostic biomarkers to aid an early diagnosis and is important for starting an early treatment of ongoing pathologies. But proteins may also serve as prognostic
biomarkers which is valuable for assessing ongoing treatment and drug development\textsuperscript{11}. Ideally, a single biomarker would be used for these purposes but usually a panel of protein biomarkers is needed to fully discriminate healthy from disease samples\textsuperscript{12}.

The path to novel clinical tools for diagnostic and prognostic purposes is a long and difficult road that may be divided into certain steps and is illustrated in Figure 2. The techniques associated with the discovery step are central to this thesis and are discussed in more detail in later chapters. A clinical proteomic study should begin with a well defined clinical question. Considerations at this initial stage should preferably involve specialists with the correct clinical knowledge. The next step is sample selection of the appropriate study populations with sufficient sample size to give enough statistical power. Individual differences (e.g. age, sex, food intake, exercise) should be considered during this stage\textsuperscript{13}. This includes also the healthy reference group which is used for the relative comparison of protein concentrations. Taking all these biological variables into consideration will minimize the inter-individual differences and facilitate finding disease-specific biomarkers. Before any analysis is performed, the samples need to be collected and stored in a proper and reproducible manner. There are many factors in this pre-analysis step which may have a dramatic impact on the results\textsuperscript{14}. For example, artifactual changes in protein expression have been reported in the analysis of samples from patients with myocardial infarction. The changes could be explained by the time interval between sample collection and analysis rather than by pathological processes\textsuperscript{15}. Factors such as site of collection, protease activity, choice of anticoagulants, sample tube, storage temperature to mention a few may ultimately contribute to the composition and integrity of the sample. Pre-analytical guidelines how to standardize these procedures are dictated from organizations such as the Clinical Laboratory Standards Institute (www.clsi.org) and the Human Proteome Organization (www.hupo.org). In general, four additional distinct phases which concern the actual analysis may be distinguished in the pipeline of clinical proteomics\textsuperscript{11}. In the discovery step, an unbiased global comparison between healthy and disease samples (~10-100s) provides input for finding significant protein alterations. The biomarker candidates that evolve from this approach are based on assessment of relative peptide or protein abundance. Some of the candidates will be false positives, especially for the lower-abundance proteins\textsuperscript{16}. All of the remaining phases are hence followed by a targeted quantitative approach specifically aiming only the biomarker candidates in order to establish which the true biomarkers are. In the verification step, the biomarker candidates are screened for their change in expression. This may be accomplished by complimentary techniques, such as enzyme-linked immunosorbert assay (ELISA) or by targeted MS approaches employing multiple reaction monitoring (MRM) of tryptic peptides\textsuperscript{17}. This step should if possible accommodate a larger number of samples (~100s). The
next step is the *validation* step which set to fully confirm observed protein biomarker identity and altered concentration. This step is a more rigorous step consisting of even larger sample sets which ideally also capture the genetic, environmental, biological and stochastic variation in the population tested (~1000s). Preferably, also different techniques and laboratories are used\textsuperscript{11}. The few biomarker candidates that pass this step may be used as clinical tools to improve patient care.

![Figure 2. Phases of clinical proteomic research.](image)

### 2.2 Analytical strategies

The proteomic strategies differ in many aspects from traditional methods for the isolation and identification of proteins. Using classical methods, one protein is usually isolated and characterized by its sequence, structure and function. In contrast, proteomics deals with the global analysis of all proteins which requires some different approaches. The principal technology that enable a global detection is MS and is the usual endpoint in a proteomic analysis\textsuperscript{18}. Protein identification involves correlating experimental data to protein sequences predicted through genome sequencing by the use of bioinformatic tools and database searching (section 6.4). The experimental data may be (i) a MS spectrum consisting of precursor ion intensities, known as peptide/protein mass fingerprinting (PMF) or (ii) a MS/MS spectrum containing amino acid sequence information (section 6.3) obtained by fragmenting an isolated protein/peptide and detecting the resulting fragment ions\textsuperscript{19, 20}. The latter approach will give much more reliable identification and is henceforth assumed. Three principal pathways may be chosen before reaching the mass spectrometer as illustrated in Figure 3. In *top-down* proteomics, separation and detection occurs on a protein level using either gel based or gel free approaches. The intact protein is fragmented in the gas phase of the mass spectrometer and the sequence is obtained from the MS/MS spectrum. In general, the larger the protein is, the more difficult fragmentation and interpretation becomes. High performance mass spectrometers, especially Fourier transform ion cyclotron resonance (FTICR-MS) with its high resolution and mass accuracy are capable of resolving more complex spectra. Despite much improvements, this technique is still limited to small proteins (<10 kDa) and is often a complement when analyz-
In *bottom-up* proteomics, separation occurs on a protein level using similar methods as previously described and are then cleaved into peptides by specific proteases. Depending on separation technique, digestion may be done either in-gel or in-solution. In general, digestion of proteins into peptides is more accepted for the routine identification of proteins in complex mixtures. The resulting peptide mixtures are chromatographically fractionated and introduced into the mass spectrometer. Identification of proteins by the analysis of their digested peptides and their corresponding MS/MS spectra has a higher degree of certainty since the method relies on several pieces of highly specific information. In the third approach termed *shotgun* proteomics, peptides are generated in the very first step. Since several peptides arise from each protein, a highly complex peptide mixture is obtained. This requires highly efficient multidimensional separation as well as a sufficiently fast mass spectrometer that is able to keep up with all the eluting peaks. Shotgun proteomics has the advantage of being gel-free as solution based methods increases the sample throughput and is usually the choice of identifying proteins in large-scale studies. The terms bottom-up and shotgun proteomics are used somewhat interchangeable in the literature although they both rely on peptide identification. Different combinations of bottom-up and shotgun proteomic methods were utilized in Paper I-V.

![Diagram of analytical strategies in proteomic research](image-url)

*Figure 3.* General overview of analytical strategies in proteomic research.
3. Biological Matrices

The biological matrices that are in focus in this doctoral thesis are cerebrospinal fluid in **Paper I** and **V** and brain tissue in **Paper II** and **III**. Both matrices are part of the central nervous system (CNS). The CNS consists of the brain and spinal cord. The CNS is shielded from the rest of the body by the blood-brain barrier

(BBB). The BBB protect the CNS from toxins circulating in the bloodstream and also maintain homeostasis inside the CNS. This is made possible by the inner walls that is mainly made up by a specialized endothelial cell layer and the surrounding tissue of the blood capillaries that severely restrict passage of circulating compounds in this region of the body

(Figure 4). The special properties of the BBB were first observed by Paul Ehrlich, who noted that intravenously injected dyes leaked out of capillaries in all regions of the body except in the CNS

. The restriction of large molecules, like Ehrlich’s dyes, to diffuse through this barrier is the basis of the BBB. Substances that do cross the brain capillaries must actively be transported through the endothelial cell membrane. Molecular entry into the CNS is very much determined by the compounds size, charge and lipophilicity

. Compounds that do not fulfill these criteria and that are needed for brain function, such as glucose are transported across the BBB by specific transporters

.

![Figure 4. The blood-brain barrier is made up of tightly spaced endothelial cells made possible by tight junctions. Astrocytes closely surround the capillary on the abluminal side by their endfeet (ef).](image)
3.1 Cerebrospinal fluid

3.1.1 Physiology

Cerebrospinal fluid (CSF) is a transparent fluid that is produced in the brain and bathes the entire CNS. It is produced by a specialized vascular structure in the central brain called the choroid plexus protruding into the lateral, third and fourth ventricles of the brain. Ependymal cells surround these networks of capillaries and form a second type of barrier that exists in the CNS, namely the blood-CSF barrier. This barrier severely restricts passage of circulating compounds but not completely. CSF is the ultrafiltrate produced as the blood plasma from the capillaries penetrates the ependymal cells into the brain. In adults, it is estimated that the total volume of the fluid is between 150-270 mL. The rate of CSF formation is approximately 20 mL/h, thus about 500 mL is produced each day. The produced fluid percolates from the interconnected spaces of the ventricular system through perforations at the fourth ventricle into the subarachnoid space (SAS). The SAS extends superficially over the whole surface of the brain and spinal cord. The CSF circulates into the SAS and down along the spinal cord and then passes upward over the hemispheres. CSF removal is then largely accomplished by absorption through the arachnoid vili along the dorsal midline of the forebrain, permitting a unidirectional flow of CSF back to the venous circulation (Figure 5a). A few important physiological functions of the CSF have been defined; (1) physical support of the CNS and shock absorbent from external force, (2) intracerebral transport of nutrients and removal of waste products and (3) control of the chemical environment.

Figure 5. a) CSF production within the choroid plexus circulates the CNS. The fluid may act as a protein repository that may be of diagnostic value and is collected either through ventricular drainage (upper) or lumbar puncture (lower). b) Distribution of the most abundant CSF proteins. Values in % correspond to the fraction of total protein content. The numbers are adopted from reference.
3.1.2 Proteome

Sampling of CSF is made possible by the procedure of lumbar puncture or by the use of ventricular drainage (Figure 5a). The composition of the CSF is characterized with very few cells (0-4 cells/µL), low protein content (0.05-0.8 mg/mL) and a salt concentration similar to blood\textsuperscript{30}. Furthermore, CSF is also a very complex sample matrix with regard to the great number of proteins present in a large dynamic range, spanning at least 10 orders of magnitude. The latter was an analytical problem that was explored by the use of HLL in \textit{Paper I}. The CSF proteome have shown little intra-individual differences from longitudinal studies and is thus under non pathological conditions fairly constant\textsuperscript{35}. The sampling procedure is important because of the protein concentration gradient along the entire spinal cord. Differences caused by the gradient may be avoided by pooling lumbar CSF collected from the same region and implies that analyses of proteins should be performed on a standardized volume of CSF, preferably the first 12 mL\textsuperscript{36}. Additionally, differences in lumbar and ventricular protein concentrations should not be disregarded. 90\% of the total protein content is represented only by about 10 proteins (Figure 5b) where human serum albumin (HSA) alone occupies 50-60\% of the total protein content\textsuperscript{34}. When comparing the most abundant proteins in CSF with the most abundant proteins in plasma, a few observations may be noticed. Firstly, some of the most abundant proteins are also highly abundant in plasma, thus the molecular flux across the blood-CSF barrier is dependent on a protein concentration gradient in the blood. Secondly, a relatively selective exclusion of larger proteins occur when the CSF is formed through filtration at the choroid plexus\textsuperscript{37}. The last observation is that although CSF may be an ultrafiltrate of plasma, it should still be considered as a matrix with its own biochemistry, because many proteins do not exist elsewhere in the body. Thus, a intracerebral synthesis of the proteins must occur\textsuperscript{38}, and it is estimated that about 20-56\% of the CSF proteins are synthesized within the brain\textsuperscript{35, 37}. This observation is of special interest because; the close contact of the CSF with the entire CNS and the equilibrium between them reflects the activity of the CNS. Neural cells shed peptides and proteins into the cerebrospinal fluid which are carried throughout the CNS. Determining the protein composition at one time point will therefore give a snapshot of the health state of the CNS\textsuperscript{39}. The CSF is hence from a bioanalytical point of view, a rich source of putative protein biomarkers for neurological and psychiatric disorders\textsuperscript{3}.

3.2 Brain tissue

The human brain is the most complex single biological structure known. In total, there are approximately $10^{11}$ neurons where each neuron connects to
10^3-10^4 synapses. Additionally, it have been estimated that the different cell types range over 10^3 making the brain a highly heterogeneous organ\textsuperscript{40}. The function of the brain depends on the complex cellular diversity and circuitry but also on the different proteins and the differential expression. The use of CPE which enables a simultaneous fractionation of hydrophilic and hydrophobic proteins was demonstrated in Paper II and III with porcine and mouse brain respectively. CPE on brain tissue may be feasible as it is generally considered that it is easier to find a biological relevant biomarker in a matrix that is closer to the actual cause of error. This matrix is hence specific and relevant for biomarker discovery. However, clinical implementation is of course hampered as it is much more invasive and this matrix is therefore restricted to biopsies in animal models and post mortem analysis in humans. For routine diagnostics, the ideal matrix would be blood as it is much more available. Biomarkers originating from the CNS will however be less suitable for biomarker discovery as they will have to pass the CSF-blood barrier and consequent dilution when diffusing into peripheral regions of the body. Brain tissue serve as a good starting point for finding new biomarkers. But for actual clinical usefulness they also need to be detectable in more available matrices. The use of CSF seems to offer a good balance between clinical implementation and usefulness of the sample. The differences between brain tissue, CSF and blood are summarized in Figure 6.

*Figure 6. Differences between brain tissue, CSF and blood for biomarker discovery.*
4. Sample preparation

When analyzing complex samples, a proper sample preparation method is essential to reduce the complexity of the sample and making it compatible with downstream separation and detection techniques. During this step the targeted analytes are extracted from the matrix in order to minimize interferences from other compounds. When analyzing proteins in a complex biological fluid such as CSF, one of the main obstacles is the high dynamic range (section 3.1.2). High abundant proteins will hamper the detection of less abundant proteins. The high abundant proteins will affect the outcome using 2D-PAGE (large overlapping spots) or LC-MS/MS (peak overlap and signal suppression) and they will also severely limit the loadability of less abundant proteins. Therefore, a proteomic sample preparation, often aim to not only enrich all proteins present in a sample, but also to decrease the dynamic range. In this chapter, the three different strategies that were used in Paper I-III and V for sample preparation are discussed.

4.1 Hexapeptide ligand library fractionation

The use of hexapeptide ligand libraries (HLL) presents a strategy of reducing the dynamic range equally for all proteins. The technique was initially introduced as a way to simplify preparation of peptides and later for the capture of proteins from complex mixtures. The basis of the method relies on millions of different affinity baits consisting of hexapeptides which theoretically have a binding capacity for every protein present in a biological fluid. The HLL is constructed by splitting a batch of porous beads into several equal parts. To each sub-batch of beads, one single type of amino acid is added and covalently attached to the bead. Each sub-batch is then mixed, washed and split again into the same number of sub-batches. For each repeated cycle, a new amino acid is added to the resulting peptide and this process continues until reaching the hexapeptide length. In this fashion all amino acids are combined in all possible combinations and each combination is present in equimolar amount. When a HLL is incubated with a biological fluid, high abundant proteins will quickly saturate their binding sites. Assuming a sufficiently large overloading of sample, less abundant proteins will instead be concentrated on the beads up until the point of saturation. Any unbound proteins are washed away, and the remaining bound proteins are eluted to yield
a sample where the overall dynamic range has been highly reduced without removing any particular proteins\textsuperscript{43}. The overall process is summarized in Figure 7. It should be noted that the binding of one particular protein does not occur with one unique peptide-protein interaction. Instead the binding of one protein occurs with various hexapeptides. A disadvantage with this technique is that the protein concentration ratios are altered which complicates quantitative studies. Theoretically this should still be possible as long as the beads are not saturated for the protein of interest.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Reduction of the dynamic range by the use of a HLL.}
\end{figure}

The ability of HLL to compress the dynamic range makes it especially interesting for samples with a high dynamic range such as CSF. This was exemplified in \textbf{Paper I} using CSF from patients with TBI. TBI is defined as the brain damage inflicted from physical force. The outcome for TBI survivors is generally poor with almost 20\% of the cases requiring hospitalization and nursing care for the rest of their life and nearly 1\% of the victims will be heavily disabled\textsuperscript{44}. As such, TBI may be described as one of the most severe and widespread public health problems in the western world in terms of mortality, morbidity and cost despite improvements in standardized programs in the neurointensive care unit\textsuperscript{45}. The primary event occurring at the time of the trauma encompasses mechanical compression at the impact zone which causes tissue deformation and hemorrhage which lead to severe neuronal damage. In the post-traumatic period, a cascade of complex biochemical processes occur, causing additional irreversible tissue damage\textsuperscript{46}. These secondary events are initiated at the time of the trauma but are taking hours to days to develop and are causing much of the irreversible damage following a trauma\textsuperscript{47}. Protein biomarkers to monitor the treatment of the injuries resulting from this post-traumatic period have been reported but still no single clinically verified biomarker exists\textsuperscript{48}. In \textbf{Paper I}, the use of a HLL was examined for protein profiling of ventricular CSF from TBI patients. HLL was
assessed for the detection of currently reported biomarkers and investigating the dynamic range. A large sample load was obtained by ventricular drainage from two TBI patients. This large sample load in combination with the compression of the dynamic range inherent with the HLL technique enabled thorough proteome identification. An estimated dynamic range of 9 orders of magnitude spanning from mg/mL to pg/mL was found by comparison of reference interval of healthy CSF of identified proteins (Figure 8). This dynamic range should be regarded as an approximation as the lowest abundant protein, the 14-3-3 protein have been reported as upregulated in TBI. In total 339 proteins (\( p \leq 0.05 \)) were identified with 130 proteins occurring in both samples. Out of the 130 overlapping proteins, 45 % were found to be associated with degenerative/regenerative processes occurring after a brain trauma. Out of the 33 protein biomarkers reported by other studies either in CSF or brain tissue, 14 were found in both TBI samples. But, out of the more established biomarkers reported by several studies, all were identified except for myelin basic protein.

![Figure 8. Comparison of reference concentration ranges in normal human CSF for 24 identified proteins using HLL.](image)

### 4.2 Cloud-point extraction

Aqueous micellar solutions of some nonionic surfactants become turbid when heated to a distinct temperature known as the cloud-point temperature. Above this temperature, the micellar solutions separate into two distinct phases. The first phase is a large volume of aqueous solution that is almost
micelle free (surfactant-depleted), the other phase is an aqueous concentrated surfactant solution\textsuperscript{49} (surfactant-rich). This phenomena may be utilized as an extraction procedure for purifying hydrophobic compounds which are solubilized inside the micelles. This temperature induced phase separation or cloud-point extraction (CPE) was initially described for the extraction of water-insoluble metal chelates\textsuperscript{50}. Later this extraction procedure was described on membrane proteins (MP) using the nonionic surfactant Triton X-114\textsuperscript{51}. MP have a hydrophobic domain which interacts with the lipid core of the phospholipid bilayer that comprises the cell membrane and makes it susceptible for CPE. Triton X-114 will solubilize the MP by replacing the phospholipid molecules and when heated above the cloud-point temperature the MP will be extracted into the surfactant-rich phase. Also, since this volume is small, the MP are concentrated. Today, solubilization of MP with subsequent CPE using Triton X-114 is a rather common procedure for purifying MP. Benefits of Triton X-114 include a advantageous cloud-point temperature of 23 °C, limiting thermal degradation and is considered to be mild, thus the hydrophobic proteins undergo minimal denaturation.

The method described in \textit{Paper II} demonstrate the first application of CPE for the simultaneous extraction and enrichment of MP and hydrophilic proteins from brain tissue for extensive proteome coverage in combination with MS. Much attention have been put on MP as they do represent a group of proteins of particular interest. This is partly due to the fact that they are important drug targets. It have for example been estimated that MP account for about 70\% of all known pharmaceutical drug targets\textsuperscript{52}. The aim of \textit{Paper II} was to develop and optimize a method to also include the hydrophilic portion that may also include clinically important proteins. The method that was developed is depicted in Figure 9. Initially, homogenized porcine brain tissue was incubated with Triton X-114 lysis buffer at 4 °C. The sample was then centrifuged to remove unwanted particulates. The temperature was then raised above the cloud-point temperature, but still below physiological temperature, followed by centrifugation to facilitate phase separation.

\textit{Figure 9.} Principle of CPE. Brain tissue is solubilized with Triton X-114 followed by a centrifugation step to remove particulates. Lastly, an increase in temperature induces the phase separation\textsuperscript{53}.
This method does not selectively enrich only proteins but also other types of compounds. As with a fatty tissue such as brain, lipids pose to be significant contaminants. Hence, different delipidation and surfactant removal protocols were also examined in Paper II. In total, five different protocol were examined and the optimal protocol found was found to be protein precipitation with tri-n-butylphosphate/acetone/methanol (1:12:1). This optimal protocol was used in combination with three different proteomic platforms. The total number of identified proteins was 184 for the hydrophobic and 205 for the hydrophilic fraction. 27 % of the proteins were found to be MP which is close to the expected value of 1/3 for which this class of proteins are encoded in the genome\textsuperscript{52}. Thus, a representative picture of the MP was obtained. Furthermore, the overlap between the CPE extracted hydrophilic and hydrophobic proteins was only 9-16 %, indicating an efficient fractionation. In conclusion, the developed method in Paper II based on CPE of brain tissue offer the possibility to monitor MP expression but at the same time also the hydrophilic portion of the brain proteome.

CPE enrich proteins and should, as opposed to HLL (section 4.1), better preserve the dynamic range in the sample. Consequently, subsequent quantitative analysis by CPE should also be possible. Moreover, to facilitate a biological interpretation, an application utilizing CPE also require that the technique is quantitative, fast and reproducible. These aspects as well as the optimal CPE method is presented in Paper III. In this paper, the optimal conditions for CPE was applied on mouse brain tissue. These conditions consisted of delipidation with tri-n-butylphosphate/acetone/methanol (1:12:1) and 1DGE-LC-MS/MS as a proteomic platform. Also, a high resolving 7T hybrid LTQ-FTICR mass spectrometer was used in this study. This rendered a deeper proteome analysis of 784 for the hydrophobic and 1234 unique proteins for the hydrophilic fraction respectively. The large increase of detected proteins employing the same search criteria compared to the previous paper are likely due to a overall better mass spectrometer and that the mouse proteome is more characterized. This distinguished proteome coverage however limit the sample throughput as the 1DGE, manual excision of gel bands and in-gel digestion is time consuming. The use of on-filter tryptic digestion similar to a previously described protocol by Wisniewski \textit{et al} was also examined in this study\textsuperscript{54}. This protocol is faster than the previously used in-solution protocol as it does not require any additional cleanup procedures. By this on-filter digestion, 425 for the hydrophobic and 561 unique proteins for the hydrophilic fraction were identified. This faster shotgun-based method still correspond to 77 % of the the proteins detected by the more elaborate method and thus give a fair insight of the mouse brain proteome and is more suitable for large-scale sample sets. Lastly, the reproducibility of the CPE extraction with sucessive delipidation and analysis was investigated by the use of iTRAQ. It was found that the RSD values were 17-22 % for the hydrophobic phase and 23-
31 % for the hydrophilic phase with average ratio values close to the expected 1:1. This is just slightly higher than the expected 20 % variation using the iTRAQ technique (Figure 10). It should be noted that the iTRAQ technique compress ratios towards unity which makes this evaluation a bit biased. This compression effect is discussed in more detail in Paper IV. Despite this bias, Paper III indicate that CPE in combination with the shotgun-based methodology is a quantitative and reproducible technique that may be used for large-scale studies of in-depth proteome analysis of brain tissue.

4.3 Immunoaffinity fractionation

Specific removal of high abundant proteins by immunoaffinity fractionation is a common approach and was used in Paper V. The technique relies on sorbents with covalently linked antibodies that are specific for some high abundant proteins. The antibodies are usually attached to Protein A/G by cross-linking to the Fc region and thus exposing the antigen-antibody binding region in a favorable fashion. After incubating the sample with the affinity media, two sets of fractions are generated; the flow through and the bound fraction. The flow through consists of all proteins that have not been targeted by the antibodies and is used in the downstream analysis. Usually a washing step is performed and pooled with the flow through fraction. The bound fraction is usually eluted by a change in pH or by use of detergents. Since only a few proteins account for the majority of the protein content (section 3.1.2), removal of these high abundant proteins will significantly increase loadability and detection capability of less abundant proteins. The overall principle is depicted in Figure 11. Most available immunoaffinity kits are packed with antibodies to remove the most abundant proteins in biological samples such as plasma or serum. This is however of less concern since
most of the abundant proteins in these types of samples are also highly abundant in CSF (section 3.1.2). In Paper V, a column consisting of antibodies for the 7 most abundant proteins was used. The reduction of the dynamic range in this study enabled global quantification of less abundant proteins to screen for biomarkers related to SCS. An alternative approach may also be used with this technique for running targeted experiments as the employed antibodies may be used for specifically enriching potential target proteins. In this approach, the target proteins are identified by analyzing the bound fraction instead. Ease of use, reusability, automation, reproducible results and a much higher detection capacity of low abundant proteins has made this technique a common choice\textsuperscript{56}. This approach is however not adoptable without considerations. Many of the columns are reusable but will over time show a loss of efficiency\textsuperscript{57}. Since HSA is a carrier protein, other non-targeted proteins may be co-depleted and the quality of the depletion is highly dependent on the specificity of the antibody\textsuperscript{58}. Pre-fractionation of CSF has also successfully been described by the use of solid phase extraction which avoids potential problems linked to the antibody\textsuperscript{59}.

*Figure 11.* Immunoaffinity fractionation relies on specific removal of high abundant proteins by the use of antibodies.
5. Separation techniques

There are many different separation modes that may be used on-line or off-line in order to fractionate and separate before introduction into the MS. The physiochemical diversity of peptides (hydrophobicity, isoelectric point, charge, size) makes them well suited to be separated by many different separation techniques. Liquid chromatography (LC) is the main chromatographic technique in proteomics and was used throughout Paper I-V. A big advantage using different modes of LC is the possible hyphenation with MS, either on-line using ESI, or off-line with subsequent MALDI-TOF analysis. Another important separation technique in proteomics is electrophoresis and is briefly discussed. Isoelectric focusing (IEF) is an electrophoretic technique that was used in combination with LC and MS detection in Paper I-II. This creates an orthogonal two dimensional (2D) setup for increased separation power. One dimensional gel electrophoresis (1DGE) was also used in Paper I-III.

5.1 Liquid chromatography

The basis of chromatography (“color writing”) was first described by the Russian botanist Michael Tswett more than 100 years ago60, 61. A chromatographic system consists of a mobile phase that carries the sample through a stationary phase. The separation of analytes in a sample occurs as they have different affinities for the stationary media. Consequently, the retention time for a particular analyte, using a given mobile phase and stationary phase, to travel through the media is dependent on its physiochemical properties. In gas chromatography (GC), the mobile phase used is a gas. The technique is suitable for low molecular compounds that are volatile and thermostable. For proteomic research this is not the case. Instead, liquid chromatography (LC) is the chromatographic technique of choice. In LC, the mobile phase is a liquid and the stationary phase usually consists of small porous particles (often silica) with a large surface area that are coated with different ligands. LC is often the method of choice for the analysis of biological samples but is also a convenient method for performing sample cleanup steps such as desalting and detergent removal. This technique is used most extensively in the form of high-performance liquid chromatography (HPLC) operating under high pressure. Main advantages of HPLC over low-pressure methods are
shorter analysis time, higher sensitivity and higher resolution. Liquid-base separations can be run in various modes defined by the different analyte-ligand interactions made possible by varying the stationary phase in the separation column. Common modes in proteomic research are ion-exchange, affinity, size-exclusion and reversed phase (RP) chromatography. It is generally considered that no single chromatographic separation is capable of resolving a complex mixture of peptides that result from a global proteolytic digest of a complex proteome. Multidimensional separations are possible to further reduce the complexity and thereby to increase the number of identified peptides and consequently increase the proteome coverage. The combination of strong cation exchange followed by RPLC is by far the most commonly used online multidimensional liquid separation approach in the field of proteomics. The 2D setup that we have adopted in Paper I-III is based on an offline peptide IEF step or 1DGE (section 5.2.1) followed by nano-RPLC and ESI-MS.

5.1.1 Reversed phase liquid chromatography

Reversed phase liquid chromatography (RPLC) deserves special attention because it was used throughout Paper I-V and it is the principal mode to be hyphenated with MS for highly efficient peptide and protein separation. The separation is based on difference in hydrophobicity of the analytes. Retention time is decreased by addition of an organic solvent, (acetonitrile/methanol) to the mobile phase. Gradient elution is generally used due to the large difference in capacity factors present in a tryptic digest. The gradient slope is an important parameter as it affects the peak capacity. In general, shallow gradients produce higher peak capacities than steeper gradients and an increase in unique peptide hits and protein identifications but on the expense of time. The peak capacity is however not the only important factor as the use of very long gradients or columns result in broad peaks with low peak heights and thus sensitivity may suffer. The hydrophobicity of the analytes is pH dependent. For this reason, ion-pairing reagents such as formic acid or trifluoroacetic acid are often added to the mobile phase which generally improves the chromatography. The choice of ligands has the greatest impact on the chromatography. The particle surface is modified with alkyl chains in varying length from C₄ to C₁₈. Chains in the range of C₄ to C₈ are mainly used for proteins and C₁₈ for separation of peptides. The choice of packing material also has large effect on the separation and resolution of proteins and peptides with regard to the particle size, pore size, surface area and silica purity. Typically, particle sizes are within the range of 3-5 µm for analytical separations. The peak capacity may be improved by reducing the particle size or by increasing the column length but both approaches are inherent with increasing back pressure. A reduction in particle diameter or
an increase in column length with a factor of 2 yields a pressure increase with a factor of 4 and 2 respectively at the same flow rate.

5.1.2 Miniaturization of liquid chromatography

One trend in LC is the reduction in the internal diameter (i.d.) of the column and adapting the chromatographic system to operate at the decreased scale. This has two major advantages compared to conventional LC. The first advantage is the lower sample requirements. For a given amount of injected sample, the analyte concentration is inversely proportional to the square of the column i.d. This decrease in column i.d. will however accompany a decrease in loading capacity. But, for a typical neuroproteomic experiment as little as only a few femtomoles of digested protein usually is available. The second advantage is the enhanced limit of detections by LC-ESI (section 6.1.1.1) Limited sample amounts and sensitivity are two major reasons, especially in the field of neuroproteomics which explains why miniaturized LC-ESI systems have been used throughout Paper I-V. The use of nano-LC represents the current technical limit by downsampling. Construction of a nano scale LC system requires some technical expertise as it places restrictions on all the extra column volumes (dead volumes) that becomes more critical when working in this scale. There are still no accepted definitions describing the different scales of LC techniques. A separation may be considered to be performed in the nano scale on the basis of the flow rate range used and that packed capillaries with a i.d. of 10-150 µm is employed. There is an overlap between the classifications found in the literature but some of the most common classifications and suggested ranges are summarized in Table I.

<table>
<thead>
<tr>
<th>Column i.d.</th>
<th>Flow rate</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2-4.6 mm</td>
<td>0.5-2.0 mL/min</td>
<td>conventional LC</td>
</tr>
<tr>
<td>1.5-3.2 mm</td>
<td>100-500 µL/min</td>
<td>microbore LC</td>
</tr>
<tr>
<td>0.5-1.5 mm</td>
<td>10-100 µL/min</td>
<td>micro LC</td>
</tr>
<tr>
<td>150-500 µm</td>
<td>1-10 µL/min</td>
<td>capillary LC</td>
</tr>
<tr>
<td>10-150 µm</td>
<td>10-1000 nL/min</td>
<td>nano LC</td>
</tr>
</tbody>
</table>

In a miniaturized LC, the optimal flow rate is lowered with the same exponential that the i.d. is reduced but in order to maintain performance all volumes need to be down-scaled accordingly. This applies to all parts of the system that may give extracolumn band broadening, such as connections, tubings, detection cells. A nano scale HPLC system is prone to clogging by particles. Hence, removal of insoluble particles by the use of spin-filters,
pre-columns or similar methods is recommended before injecting the sample into the system. This will also extend the lifetime of the nano HPLC column. Some of the injections that were made in a nano LC system in Paper I-IV were carried out by the use of a RP-trap column. By this method, an injection volume of a few µl is performed using a low organic content, stacking the peptides onto the low-volume trap column. After loading the peptides onto the trap column an external valve switches and the peptides are eluted onto the nano HPLC column. Flow rates in the µL/min range can be used during the loading step onto the trap. The size of the trap should be carefully selected in order to provide a sufficient loading capacity whilst maintaining separation efficiency. In order to reduce the extra column band broadening the connections need to be done in a way that minimizes the dead volumes. For the connections of a nano LC system, usually zero-dead volume unions in which the tubing ends are closely adjoined to each other are used. This also puts restrictions to how the tubing is cut. When using ESI-MS, the spray needle and the column may be incorporated into one entity by packing the stationary phase directly into the spray needle. This creates a fritless column with minimal need for connections and extra dead volumes.

5.2 Electrophoresis

Electric field driven separations are also common in proteomics and many other fields. An applied electric field causes ions to travel through different medias such as a gas, liquid or gel. The rate of movement is dependent on properties of the analyte such as size, charge and shape but also the field strength and media. Gel based separations employing matrices such as polyacrylamide or agarose have historically had an important role in the separation of proteins and peptides. Analysis using MS is possible by the use of in-gel digestion. However, this approach is time and labor consuming. Since MS is the main detector in proteomics, the trend is moving towards more liquid based separations that may be hyphenated more directly with MS.

5.2.1 Isoelectric focusing

Isoelectric focusing (IEF) separates amphoteric analytes such as proteins or peptides on the basis of charge as defined by the pKₐ values of proton accepting sites. For proteins and peptides, these sites are found in side chains of acidic and basic amino acids as well as the N- and C-termini. The isoelectric point (pI) is defined as the pH where an amphoteric molecule is in a net neutral charge state. The basis of IEF is to use a separation media consisting of a pH gradient that encompasses the pI values for the analytes of interest. The acidic end of the gradient is placed to a positive electrode and the basic end to a negative electrode. Upon applying an electric potential the analytes
will start to migrate towards the steady state position corresponding to its pI where the counteracting forces from the electrodes are zero. The use of immobilized pH gradients (IPG) in a polyacrylamide gel is often the preferred choice. The major benefit being that the gradient is fixed to the matrix. The 2D setup that was used throughout Paper I and II employs a IEF step as first dimension using the Off-Gel 3100™ system. The Off-Gel system is a multi-compartment separation system where the sample is placed in liquid chambers that are positioned on top of an IPG gel (Figure 12). Thus, separation partly takes place in the gel but has the main advantage of an in solution recovery of the analytes which makes it more compatible for MS.

![Image of principle of the Off-Gel 3100™ system](image)

**Figure 12.** Principle of the Off-Gel 3100™ system. a) The sample is distributed among the micro-wells. b) High voltage is turned on and the analytes will migrate along the pH gradient towards the position corresponding to its pI value. c) Once at a steady-state, a low potential keep the analytes in place until harvesting.

The offline IEF is more time consuming and do involve some additional steps that are associated with potential sample loss compared to an online multidimensional method. Advantages with this offline IEF as 1:st dimension are (1) high start loading capacity, (2) high resolving power, (3) ease of RP gradient implementation as no additional LC pump is needed.

### 5.2.2 Gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is one of the most widely used electrophoretic techniques as it is simple, fast and allows an easy visual interpretation of the separated proteins. The separation mechanism is based primarily on mass but also size and shape influence the separation. SDS is an effective solubilizing detergent with high affinity for proteins. In a protein mixture, SDS binds to proteins in numbers proportional to the protein mass and each SDS contributes with one negative charge. Consequently the number of charges will be proportional to the size of a protein. It also promotes protein denaturation but in order to achieve complete unfolding, the addition of a reducing agent to cleave disulfide bonds is needed.
bonds and an alkylating agent to prevent reoxidation of the disulfide is done before performing SDS-PAGE. When applying voltage, all SDS-protein micelles will migrate towards the anode and their electrophoretic mobilities will mostly depend on the size of the protein. Smaller proteins will migrate faster due to a higher electrophoretic mobility but also because they will migrate faster through the pores in the acrylamide gel. After fixing the proteins, visualization of the proteins may be obtained by different staining methods. Silver staining has a detection limit below 1 ng but is usually not MS compatible and has a limited linear range of only one order of magnitude. Coomassie brilliant blue is a cheap and MS compatible dye that has a high linear range of up to three orders of magnitude but has a disadvantage with a detection limit of only about 10 ng. Another common dye is SYPRO Ruby that combines all the advantages already mentioned, but requires a fluorescence detector. The use of SDS-PAGE with coomassie staining was used in Paper I as a screening method to check for protein enrichment, in Paper II for finding the optimal delipidation procedure and in Paper II-III also as a pre-fractionation step.

A significant improvement of SDS-PAGE is the use of two-dimensional electrophoresis (2DGE) as first described in 1975. This technique has not been utilized in any of the included papers but deserves special attention because it remains one of the core tools in expression proteomics. This is mainly because of its very high resolving power, capable of resolving up to several thousands of proteins simultaneously on a single gel. 2DGE consists of combining the electrophoretic methods already described with IEF in the 1:st dimension and SDS-PAGE in the 2:nd dimension. Commercially available analysis software exploits comparing protein pattern images in a digital format obtained from the use of scanners, phosphor imagers or other type depending on chosen staining procedure. The proteins will not migrate to the exact same point between gels but a software correction by matching of “landmark proteins” will compensate for any shifts and improve gel to gel comparisons of individual spots. Automated spot detection and gel alignment algorithms facilitate the throughput but a manual user intervention is often necessary to compensate for any false matches. It is however not a technique without its limitations; detection of low abundant, very small or large proteins, as well as basic and hydrophobic proteins is still problematic. The probably main disadvantage is that despite software with automated, gel alignment and spot picking it is a time consuming approach not easily adoptable for high throughput studies.
6. Mass spectrometry

Mass spectrometry (MS) is an important tool in many fields of science but for the characterization of individual isolated proteins or on a proteomic-wide scale it is an indispensable technique. The evolution of MS started in the beginning of the 19th century. Since then, technical developments have made it possible to ionize and analyze more classes of compounds. The breakthrough invention of electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) made it possible to ionize nonvolatile, high-molecular mass compounds and was awarded the Nobel Prize in 2002. ESI and MALDI are the dominating methods for peptide and protein analysis today. A mass spectrometer consists of three major parts, an ion source, a mass analyzer and a detector. The role of the ion source is to create gas phase ions, either continuously from a flow of liquid as in the case of ESI or as a discrete process as MALDI. The mass analyzer is capable of separating the ions according to their mass to charge ratio (m/z). This manipulation of ions is only possible without interferences from other gas molecules; hence low pressure is maintained by a pumping system. The final part of the mass spectrometer is the detector. Its role is to convert the energy of an incoming ion into a current signal that is registered by the acquisition computer. This chapter will mainly focus on ESI-MS as it was used throughout Paper I-V. Since MALDI-MS was used in Paper II to some extent and because of its importance it will also be discussed.

6.1 Ionization techniques

6.1.1 Electrospray ionization

The ESI process was after its invention combined with MS by Fenn and coworkers in 1984. The sample is dissolved in a polar, volatile solvent and transferred through a capillary (needle). When a high electric potential (1-5 kV) is applied between the capillary and a counter electrode, the liquid will at the tip of the capillary form a cone shaped formation known as the Taylor cone. An aerosol spray consisting of fine charged droplets are ejected from the Taylor cone which is enriched with either positive or negative ions. The basis of this phenomenon will be further explained in the positive ion mode
where a positive potential is applied to the tip as in Figure 13. Ions in the solution will experience an electrophoretic migration as cations will migrate towards the surface of the liquid and anions will migrate inwards the capillary. When the repulsion between the cations at the surface and the electrostatic attraction towards the counter electrode overcome the surface tension, the surface of the liquid elongates into the Taylor cone. At a certain voltage (onset voltage), the established electric field exceeds the surface tension and the Taylor cone breaks up into a liquid jet where small positively charged droplets are ejected. The droplets will shrink as the ratio of surface charge to area will increase during its path to the counter electrode and finally end up as desolvated ions. This evaporation process is usually aided by a pneumatically flow of nitrogen gas along the outer walls of the capillary and heat. The generated ions are thus formed at atmospheric pressure and enter the mass spectrometer through a small orifice into an intermediate vacuum region and from there through a small aperture into the high vacuum of the mass analyzer. Successful combination of ESI with virtually any analyzer is possible.

![Figure 13. Formation of ions by the electrospray process.](image)

A prerequisite for ESI is that the analytes are able to form ions in solution. For this reason RP is a compatible separation mode before ESI (section 5.1.1), where the mobile phase composition is usually a mixture of water and organic solvent such as methanol or acetonitrile. A small content (0.1-1 %) of formic or acetic acid will aid protonation in positive mode, whereas ammonia will aid deprotonation in negative mode of the analytes. ESI-MS is suitable for analysis of polar molecules covering a wide mass range from about 50 Da up to large proteins and even a whole virus with a mass of more than 2 MDa\textsuperscript{87}. Analysis of large molecules is possible even with mass analyzers with limited m/z due to the capability of ESI to generate multiply charged ions. Changes in the analysis conditions such as solvent composition, pH and partial denaturation will alter the charge state distribution of large molecules\textsuperscript{88}. The complexity in the spectrum due to multiple-charged
ions will make interpretation of ESI from complex samples difficult. This further emphasizes the necessity of having a good separation before ESI detection. Another disadvantage of multiple charge states is that the signal resulting from an analyte is distributed over multiple signals which decrease the sensitivity. For tryptic peptides however, most peaks are either only doubly or triply charged. ESI is a soft ionization method, accompanied by very little fragmentation of the formed gas phase ions. Consequently, weak bonds are often preserved and analysis of intact post-translational modification and non-covalent complexes is possible with ESI-MS89.

### 6.1.1.1 Nano electrospray ionization

The use of nano-ESI is especially suited for neuroproteomics mainly due to low sample consumption and low detection limits and has been used throughout Paper I-V. The reason is that in conventional ESI, the response of an analyte by continuous infusion will give a constant signal which is independent of the flow rate and is only dependant of the concentration (concentration sensitive). This is true with flow rates using conventional LC separations and tip geometries of the capillaries used for regular ESI. In this range only a small amount (about 1 %) of the introduced analytes are ionized and the diameter of the initial droplets are in the µm range90. However, if the flow rate is lowered into the nL/min range in parallel with using a capillary tip with reduced aperture, a change in the response behavior will start to appear. Below a certain flow rate, the ESI-MS response will start to decrease (mass sensitive). This will occur when the efficiency of the conversion of liquid phase ions to gas phase ions (ionization efficiency) remains constant as the flow changes. The ionization efficiency has reached its maximum at the point where signal start to decrease. Thus, maximum utilization of a sample accompanied with maximum sensitivity may be obtained at the flowrate just above this point91. These desirable properties of nano-ESI may be obtained by using flowrates in the nL/min range and employing needles with smaller i.d. and reduced aperture92, 93. This mode of ESI has higher ionization efficiency because of the reduced initial diameter of the droplets that are ejected from the Taylor cone. These smaller initial droplets are in the nm range and require fewer fission events and less solvent evaporation prior to release of gas phase ions that in together result in faster ion emissions94. Nano-ESI tips may thus be placed in close proximity to the entrance aperture of the mass spectrometer, maximizing the collection of the ion beam without presenting the vacuum system with jets of liquid. Nano-ESI is also less prone to suppression effects and is thus more tolerable to salt than conventional ESI92. Additional differences between conventional ESI and nano-ESI are summarized in table II.
Online nano-ESI is best used in conjunction with nano-LC (section 5.1.2) for pre-fractionation of a complex sample and presents a fruitful merger. However, the technique also has disadvantages. The usage of small i.d. of the analytical column and the emitter tip are prone to clogging. Reproducibility of producing the narrow tips is difficult and they are expensive. The robustness of the spray is poor and requires a skilled operator along with long analysis time. All these disadvantages have limited the widespread use of the technique and is in practice only restricted to applications with limited sample amounts that require lower detection limits.

6.1.2 Matrix assisted laser desorption ionization

MALDI was developed by Karas and Hillenkamp\textsuperscript{95, 96} as a new ionization method for the analysis of peptides and proteins. This technique is a sensitive and soft ionization technique suitable for large intact molecular compounds. In the MALDI ionization process, a low molecular light absorbing matrix is added in excess and is co-crystallized with the sample on a surface. Upon irradiation of the crystallized spot by an ultraviolet (or infrared) laser pulse, the analytes are desorbed from the surface into the gas phase and ionize by a still unknown mechanism as shown in Figure 14. The ionization method produces mostly singly charged ions which makes the interpretation of the resulting spectrum simpler than an ESI spectrum.

\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
\textbf{Parameter} & \textbf{ESI} & \textbf{Nano-ESI} \\
\hline
Flow rate & 1-1000 \( \mu \text{L/min} \) & 1-1000 \( \text{nL/min} \) \\
Voltage & 3-5 \( \text{kV} \) & 1-3 \( \text{kV} \) \\
Sample consumption & pmol to fmol & fmol \\
Sprayer diameter & 50-200 \( \mu \text{m} \) & 1-25 \( \mu \text{m} \) \\
Sensitivity & \text{~fmol} & \text{~amol} \\
Nebulization gas / Heat & Necessary & Optional \\
\hline
\end{tabular}
\caption{General differences between conventional ESI and nano-ESI}
\end{table}

*Figure 14.* Irradiation by a laser will cause the matrix and analyte molecules to desorb into the gas phase. In the gas cloud a charge transfer takes place that will produce mostly singly charged ions.
Despite lack of complete understanding of the mechanism, the matrix has a central role in the process of absorbing energy and promoting proton transfer. The choice of matrix is crucial for successful analysis and is chosen on the basis of analyte properties. A wide range of matrices exist. Examples of popular matrices are α-cyano-4-hydroxycinnamic acid which is suitable for peptides up to 5 kDa and sinapinic acid which is suitable for proteins. A common property of the matrices is that they are UV absorbing with delocalized π-electrons. Several techniques for sample and matrix deposition on the target plate exist. A popular method is the dried-droplet method where a drop of aqueous matrix solution is mixed with the sample and dried at room temperature. After dryness, sample cleanup is possible by immersing the plate in a washing solution before measurement. However, MALDI is known to be less sensitive than ESI to ion suppression and thus tolerate salt and additive contamination to a higher extent. The method employed in Paper II used an AnchorChip™ plate consisting of 384 pre-spotted dots with matrix. The tryptic peptides were separated with RPLC and the eluate was deposited directly onto the spots by the use of a MALDI fraction collector for automation. Regardless of the deposition technique, the off-line character of MALDI is more time consuming but is also advantageous since reanalysis of the sample is possible.

6.2 Mass analyzers

Ionized gas phase peptides are introduced into the mass spectrometer by the ion source and are then analyzed by the mass analyzer. Identification of as many proteins/peptides as possible is usually performed by operating the mass analyzer in information dependant acquisition (IDA). This is usually performed by an initial precursor ion scan, wherein each intact peptide ion produces a peak in the mass spectrum. Usually, the mass spectrometer then dynamically selects one or more of the peaks to be isolated and fragmented. The mass spectrum acquired for the fragment ions from the initial peptide ion is the MS/MS scan. This cycle of an initial precursor ion scan followed by several MS/MS scan are repeated throughout the analysis. The performance of the mass analyzer performing these tasks is critical for the quality of the data. The ideal mass analyzer should have high resolution, mass accuracy, mass range, ion transmission, dynamic range, speed and robustness at a low cost. However, a mass analyzer that possess all these qualities does not exist and consequently several types of mass analyzers exist; each with its advantages and disadvantages. MS/MS may be achieved by (i) combing two mass analyzers with an intermediate step of fragmentation (tandem-in-space) or (ii) within one and the same mass analyzer as with trapping instruments where the various stages of MS/MS are conducted at different times during the experiment (tandem-in-time)\textsuperscript{97}. Also, the combination of different mass
analyzers (hybrids) may be combined to provide different performance characteristics.

6.2.1 Quadrupole
The quadrupole (Q) consist of a set of four parallel rods arranged symmetrically and employ a combination of direct-current (DC) and radio frequency (RF) potentials. Ions are accelerated and introduced into the Q in the center of the rods. At a given value of DC and RF potential, only ions within a certain m/z range will have stable trajectories through the Q (resonant ions). All ions outside this m/z range (non-resonant) will have an unstable trajectory and collide with the rods. A mass spectrum is acquired by varying the DC/RF ratio and thus scanning through the m/z range of interest. A Q has a rather low mass accuracy and resolution, usually unit resolution of 0.7 Da at full width at half maximum. The Q is instead chosen for its speed, high dynamic range, simplicity in maintaining and sensitivity. A popular combination is the triple Q instrument (QqQ) that allows MS/MS. It consists of two stages of mass filtering Q with an intermediate Q serving as a collision cell. Low cost, robustness and multiple scan techniques is associated with QqQ instruments. Of the different scan techniques, multiple reaction monitoring (MRM) is the most important mode for targeted analysis. In this mode, both mass filters of the QqQ instrument are static for a given set of precursor ions and corresponding fragments. This mode will provide the highest sensitivity. The use of hybrid instruments using Q’s in different combinations were used throughout Paper I-IV.

6.2.2 Quadrupole ion trap
A quadrupole ion trap (QIT) is based on the same principles as the Q but has a different geometry that allows ions to be trapped inside a potential well. Typically the well consists of one ring electrode and two cap electrodes with openings for injection and ejection of ions. Trapping of ions may be achieved by applying a low RF potential to the ring electrode. As the RF potential is raised, ions of ascending m/z will become unstable and are gradually ejected and thus producing a mass spectrum (mass selective instability scan). Isolation of ions with only a selected m/z is also possible (isolation scan). Combination of successive isolation scans and fragmentation allow MS^n for structure elucidation. The resolution and mass accuracy is in general better than for a Q but are not considered to be high. The strengths of the QIT are fast scan rates and low limit of detections. The latter partly due to the trapping capability that may accumulate a sufficient ion package for reaching lower detection limits. The dynamic range is rather limited for a QIT. The reason for this is that it can only hold a certain number of ions before repulsive forces (space charging) may arise. The use of linear ion
traps (LIT) improves this limitation. The LIT is built like a Q but has two electrodes for axial trapping similar to the cap electrodes of the QIT. The LIT has a larger volume than the QIT and can thus store more ions before space-charge effects appear. The QTrap used in Paper I-II was a hybrid consisting of a Q with a collision cell and a LIT.

6.2.3 Fourier transform ion cyclotron resonance

As with a QIT, a Fourier transform ion cyclotron resonance (FTICR) analyzer is capable of trapping ions. This is achieved by using a strong magnetic field that will cause the ions to move in a circular motion around the ICR cell. The cell is usually cylindrical and consists of opposing plates for trapping, excitation and detection and is placed inside the magnetic field. The circulating ions will move around the cell at a frequency that is specific to their m/z ratio. This frequency is known as the cyclotron frequency. To detect the ions, a RF voltage is applied to the excitation electrodes that match the cyclotron frequency. This will cause the ions of a certain m/z ratio to be excited to a larger orbital radius. The excited ions will approach the detection plates and induce a small current for each cycle. To acquire a mass spectrum, a RF voltage that matches all ions in a certain mass range is applied. This will cause the ions within the m/z range of interest to be excited to a larger orbital radius and simultaneous detection by the detection plates. This time domain signal may then be processed by a Fourier transformation into a mass spectrum. The advantage with a FTICR is the unparalleled resolution that may be obtained. This analyzer was used in Paper III-V and consisted of a LIT followed by a FTICR. This hybrid combines the speed and sensitivity associated with the LIT and the high resolution of the FTICR. The handling of the instrument is however demanding due to cooling by liquid nitrogen and helium for maintaining the superconductivity of the magnet.

6.2.4 Time-of-flight

Due to the pulsed nature of MALDI, the time-of-flight (TOF) is the most often used mass analyzer in conjunction with this ion source. This combination was also used in Paper II. Parameters of modern TOF analyzer provide high resolution, high accuracy, high mass range (in theory no limit), high speed (~µs/pulse) and sensitivity. In the case of MALDI, the ions enter the gas phase and are accelerated towards the detector. When leaving the acceleration region, all intact ions will ideally have the same kinetic energy but different mass-dependent velocities. The ions are then allowed to drift in a field-free region towards the detector. The time difference for the ions between entering the field-free region and hitting the detector is the time-of-flight. Lighter ions will have faster velocities than heavier ions and the TOF spectrum can be converted into a mass spectrum. Another common hybrid
employing TOF analyzers are QTOF instruments. This combination was used in the QStar used in Paper III-IV.

6.3 Fragmentation and sequencing of peptides

Fragmentation of peptides may be performed by collisions of an inert gas (N₂ or He) termed collision induced dissociation (CID). The amino acid sequence may thereafter be determined by interpreting the product ion spectrum resulting from MS/MS experiments. This may also be done on a protein level but digestion of proteins into peptides presents a class of smaller compounds with bonds of similar chemical stability where fragmentation and interpretation may be handled. MS/MS is usually performed on tryptic peptides as its specificity is well suited for MS/MS. These peptides will have an arginyl or lysyl residue at the C-terminus that will readily incorporate a proton as well as a basic N-terminus. Using positive ESI at low pH, tryptic peptides are thus at least doubly charged which improve ionization yield. Additionally, trypsin generates peptides of suitable length for MS/MS. As shown in Figure 15, several bonds may be broken along the peptide backbone. In CID, the most common ion types are the b and y ions which denote fragmentation at the amide bond with charge retention on the N or C terminus respectively. Ideally, an MS/MS spectrum of a peptide will provide a ladder of peaks where the distances between them reflect the sequence of amino acid residues. Upon using CID, the sequence is obtained by reading either the y or b ion series from left to right in the spectrum. In practice however, complete series are rarely found and parts of a sequence is interpreted with confidence. Additional sequencing information may be obtained as almost every bond in a peptide may be broken upon fragmentation.

**Figure 15.** The Roepstorff’s nomenclature of peptide ions resulting from backbone fragmentation.
6.4 Bioinformatics

Manual interpretation of data obtained from a MS/MS experiment of peptides for protein identification is not feasible in a high-throughput manner. The use of computer assisted identification and validation of the data is therefore necessary. A large number of computational methods, search engines and databases exist to automate this task from MS or MS/MS data. Three main approaches are used: (i) *de novo* sequencing utilizes solely the information from an MS/MS spectrum and attempts to read the sequence directly from ion series and fragmentation rules; (ii) database searching by matching the experimental data against spectral libraries consisting of reference spectra with verified sequences or; (iii) matching against theoretical spectra predicted for each peptide contained in a protein database. Regardless of the method, they are all probabilistic methods which generate a result associated with a certain degree of probability.

In *Paper I-V*, the search engine Mascot (www.matrixscience.com) has been used with the last approach. Before the Mascot search, user defined criteria about the acquired experimental data is defined to increase the specificity. These criteria include the biology (e.g. taxonomy) and modifications (e.g. protease) MS settings (peptide tolerance) and also the preferred database. The data is submitted in the form of peak lists with the precursor ion and associated fragment ions from the MS/MS spectrum including their intensities. Each MS/MS spectrum is searched individually by using a specific search algorithm that computes the probability that the observed match is a random event. The correct match, which is not a random event, consequently has a low probability. This probability score is then converted into a peptide ion score in which a higher value means a higher degree of similarity between the experimental spectrum and the theoretical spectrum. The output is a list of proteins ranked according to their protein ion score which is based on all the individual peptide ion scores. For a peptide to be considered to be a hit, the peptide ion score must pass a significance threshold which is usually set to at least 95 % ($p \leq 0.05$) or higher. The absolute value of the significance threshold depend on numerous factors as for example the size of the database or the number of queries. Only peptide matches with ion scores above the significance threshold may be chosen to contribute to the total protein ion score (MudPit scoring). This filtering will limit false positive protein matches due to accumulation of low scoring peptides and have been used throughout *Paper I-V*. An additional validation step of the data is the target-decoy strategy. It involves researching the data against a database consisting of decoy peptide sequences from the original database but in one way manipulated (scrambled or reversed). Assuming that matches to decoy peptide sequences and false matches to sequences from the target database follow the same distribution, then the false discovery rate may be calculated.
This value does in percentage tell how many proteins in the first search may be considered to be true positive hits.

6.5 Protein Quantification by Mass Spectrometry

The methodological and technical developments that have been discussed in the previous chapters, from the initial sample preparation to identification by MS and bioinformatics, have enabled global protein identification in complex biological samples. This data need to accompany quantitative information in order to fully understand the roles of proteins in an organism. A comparison of a biological sample (disease) needs to be performed to another reference sample (healthy). This way of detecting protein differences allow a judgment whether a certain protein should be investigated further as a potential biomarker.

In comparative proteomics, this may be achieved by label-free (LF) methods by comparing the inter-run MS peak intensity for any given peptide or by using the number of acquired peptide spectra as an indicator for the amount of a protein\textsuperscript{103}. More common is the use of label-based techniques where stable heavy isotopes are incorporated in order to generate a specific mass tag\textsuperscript{104}. Peptides from the biological sample are derivatized by a certain “light” label and the reference sample by a “heavy” label (or vice versa). This way of multiplexing allow for comparing the intra-run MS peak intensities. The advantages by using labels are reduced analysis time, higher sample throughput, less experimental variation due to alignment of peaks and simplified quantification. The resulting spectrum will show pairs of peaks that are separated due to the mass difference and the intensity of the peaks provides the basis for the relative quantification. Dimethyl labeling (DML) is a labeling technique for up to three samples that perform quantification in this manner\textsuperscript{105, 106}. Upon this labeling technique, dimethyl groups are incorporated to lysine and peptide N-termini which gives a net mass shift of 4 Da between the “light” “medium and “heavy” label respectively. The overall principle is depicted in Figure 16a.

A refined way of chemical derivatization is to use isobaric tags. Isobaric tag for relative and absolute quantification (iTRAQ) allow 4-8 plex labeling\textsuperscript{107, 108}. The reagents consist of a reporter, balancer and a peptide reactive group. In the iTRAQ 4-plex, the overall mass of the reagents are kept constant using a combination of different isotopic elements of the reporter and balancer group such that the combined mass is 145 Da. The peptide reactive group specifically forms an amide bond to N-termini and lysine residues and thus adds a mass shift of 145 Da to tryptic peptides. By a MS survey scan, a pool of four iTRAQ labeled peptides will appear as a single peak with the intensity from each of the peptides adding to the total peak intensity. Upon CID, the four reporter ion groups will appear as distinct peaks in the low
mass region. The quantification may then be achieved in the MS/MS mode by using the intensity from each of the reporter ion groups as shown in Figure 16b.

Figure 16. Protein quantification by chemical derivatization (a) DML allows 3-plexing and utilizes isotopomers of formaldehyde and sodium cyanoborohydride to incorporate dimethyl groups to proteolytic peptides. A mass difference of 4 Da allow the peptides to be distinguished and quantified by MS. (b) iTRAQ allow 4-plex and adds an isobaric mass of 145 Da to proteolytic peptides. A single peak will appear from a mixture of four iTRAQ labeled peptides in MS mode. Upon MS/MS, the reporter ion groups are released and may be used for distinguishing the peptides and quantification.

The necessity of understanding the advantages but also the limits of a method are crucial in order to choose the best method. A comparison between iTRAQ, DML and LF for relative protein quantification using shotgun proteomics was the main focus in Paper IV. The comparison was based on experiments with (i) samples having a few differentially expressed bovine proteins and (ii) samples having the same differentially expressed bovine proteins and a bulk of E. coli proteins occurring in unchanged ratio. The latter two proteome model will mimic the conditions in an actual biological sample where most of the matrix proteins are expected to occur in a 1:1 ratio and a few proteins that may be regarded as biomarkers. Also, this type of experimental design will better illustrate the performance of each method than by comparing protein ratios between samples where all of the proteins in each sample have changed with a given ratio. Moreover, the performances of the methods in Paper IV were investigated using two different mass spectrometers, a QTOF and a high resolving 7 T FTICR.

LF FTICR was found to by far have the highest proteome coverage (95%) of the matrix proteins. This was followed by iTRAQ (47%)>DML FTICR (46%)>LF QTOF (41%). DML in combination with the QTOF gen-
erate a substantially less proteome coverage (14%). The highest accuracy based on the artificially regulated proteins was found for DML FTICR (54%). Overall better results were obtained with the DML in combination with the FTICR because of the faster employed duty cycle that is required with the increased spectrum complexity inherent with DML. All methods were found to possess acceptable linearity within the investigated dynamic range. However, all methods were found to consistently underestimate protein ratios in the presence of matrix proteins occurring in a 1:1 ratio. These underestimations of regulations are likely to arise from co-eluting isobaric ions from the matrix proteins that compress the ratios towards unity. It was confirmed in Paper IV that iTRAQ is prone to this compression effect which have been previously reported\textsuperscript{109}. LF FTICR was found to be more tolerant towards this compression effect.

The ability to quantify regulations is the main purpose in comparative proteomics but the ability to find unregulated proteins and also classify them as actually being unregulated is important to avoid false positive biomarkers. This was investigated by using the false quantification rate (FQR), defined as the number of wrongly quantified proteins divided by the total number of quantified proteins expressed in percentage. This largely depends on the threshold and the number of peptides. Paper IV also illustrates how the FQR varies at different threshold and depending on how many peptides were used for the protein quantification. It was found that iTRAQ performed acceptable with >90% correct ratios at a 50% threshold even based on single peptide quantifications. But, it should be noted that the compression effect of iTRAQ will favour this type of evaluation. For DML and LF, five and six peptides respectively are recommended to insure a reliable result. This evaluation considers only the point estimate and does not include the variation for each protein. This is therefore a rough estimation but may still serve as a guideline.

In order to get an overall assessment of the different methods, a ranking system was utilized. Several important analytical parameters such as accuracy, precision and linearity were included in the evaluation. But the evaluation was also based on properties specifically important for proteomic research such as required number of peptides, false quantification rate and proteome coverage. By this ranking system, the overall performance of the five different methods were; DML FTICR > iTRAQ QTOF > LF FTICR > DML QTOF > LF QTOF. Thus, in Paper IV it was found that the best method was DML FTICR. This setup was therefore used in Paper V.
7. Biomarkers for Spinal Cord Stimulation

7.1 Background

The sensation of pain is inherent with all higher vertebrates and is a fundamental mechanism to prevent damage and ultimately to promote survival. The sensation of acute pain, arise from the activation of primary nociceptive afferents and processing within the nociceptive system and have a protective and physiological important purpose. A different type of pain which in contrast to nociceptive pain, have a maladaptive response to an injury is neuropathic pain. This type of pain may arise by activity generated within the nociceptive system without adequate stimulation. The International Association for the Study of Pain defines neuropathic pain as “pain caused by a lesion or disease of the somatosensory nervous system”\textsuperscript{110}. Existing pharmacologic treatments for neuropathic pain is limited, with no more than 40–60\% of patients obtaining partial relief of their pain\textsuperscript{111}. An alternative treatment when pharmacological agents fail is the use of spinal cord stimulation (SCS), which has proven to be effective both in terms of efficacy and cost-effectiveness for neuropathic pain\textsuperscript{4}. Its widespread use is demonstrated by the fact that about 18 000 new implants are inserted every year\textsuperscript{112}. Clinically, SCS is accomplished by placing an electrode in the dorsal epidural space in close proximity of the affected spinal segments. Electrical pulses are emitted towards the spinal structures to elicit an analgesic effect and the duration is controlled by the patient using a remote control. Although the clinical benefit of SCS is unmistakable, a detailed knowledge of the neurochemical mechanisms that mediate the analgesic effect is still missing. The fragmentarily evidence that exist have been performed in animals which have shown that SCS-induced analgesia appears to correlate with increased spinal levels of the neurotransmitters acetylcholine (ACh) and \( \gamma \)-aminobutyric acid and with reduced glutamate concentration\textsuperscript{113, 114}. A lot of evidence also indicates that the analgesic effect achieved by SCS is mediated through the dorsal column. Also, in the dorsal horn of the spinal cord the well known gate theory of pain acts as a specific pain modulator system\textsuperscript{115}. However, this theory is not able to fully explain how SCS does generate an analgesic effect.
7.2 Proteomic profiling of Spinal Cord Stimulation

In Paper V, a proteomic assumption-free approach was used to determine which proteins might be involved in the SCS mechanism. The novelty of this paper mainly lies in; (1) it is the first proteomic approach to study the effect of SCS (2) it has never been carried out in humans before and (3) each patient is its own reference which will limit any inter-individual differences. 12 voluntary patients were asked to turn the stimulator off for a period of 48 hours before CSF collection. The patients were then asked to use the stimulator as they would in their every day routine for three weeks and a second CSF sample was collected. This experimental design allowed the collection of two unique samples; one where pain is experienced by the patient (denoted off) and another where the pain is inhibited by SCS (denoted on). This repeated sampling from each patient facilitated a relative quantitative comparison of the CSF proteomes. The proteomic platform initially employed immunoaffinity fractionation (section 4.3) to remove the seven most abundant proteins. The samples were labeled by DML with a light (off) and medium (on) label (section 6.5). The labeled and pooled samples were then chromatographically separated by RP-nanoLC (section 5.1.2) and ionized with nano-ESI (section 6.1.1.1) to maximize the use of the limited protein amount available. The multiplexed samples were analyzed with a FTICR mass spectrometer operating in IDA mode. As a quality control of the employed methodology, a pool of unpaired CSF samples was created. These quality controls (QC) had an expected 1:1 ratio and were processed together with the patient samples.

7.3 Results and discussion

The FQR for all the proteins quantified with two or more peptides was found to be merely 4 % in the QC. Thus, 96 % of the protein ratios from the patient samples may be considered to be correct. In total, 419 unique human proteins could be quantified in the patient material. Proteins that could be quantified in at least half of the patients and by at least two peptides were subjected to two stringent statistical tests. The vast majority of the proteins were found to be unchanged in response to SCS. However, seven proteins were found to be statistically differentially regulated upon the use of SCS. Five proteins were found to be up-regulated; Lynx1 (P=0.000048), Klk6 (P=0.00058), Angt (P=0.00057), A4 (P=0.0052) and Sap3 (P=0.0076). Also, two proteins were found to be down-regulated; Co2 (P=0.0046) and Ibp6 (P=0.0071). Out of these seven proteins, the most significant (P=0.000048) change was found for Lynx1. It also showed the average largest (59%) and most consistent change in all patients (11 out of 12). This protein is an endogenous toxin-like modulator of nicotinic acetylcholine receptors.
(nAChRs) activity in the CNS\textsuperscript{116}. Also, affinity towards muscarinic AChRs has been reported\textsuperscript{117}. It is structurally and functionally similar to snake venom \(\alpha\)-neurotoxins. Like \(\alpha\)-neurotoxins, it contains 10 conserved cysteine residues that give it a characteristic three finger fold that is essential for binding. Furthermore, Lynx1 is a membrane protein that is tethered to the cell plasma membrane by a glycosylphosphatidylinositol (GPI) anchor that keeps it close to the point of activity\textsuperscript{117}. Interestingly, the protein Sap3 also found to be up-regulated in our panel of seven differentially regulated proteins has a reported potential phospholipase activity\textsuperscript{118}. Thus, Sap3 has the potential to hydrolyze the GPI anchor and detach Lynx1 from the cell membrane.

The mechanism of SCS-induced analgesia is as mentioned in the foregoing discussion only fragmentarily known. However, the cholinergic system has been shown to be involved in the pain relieving effect of SCS as increased spinal levels of ACh in a rat model of neuropathy has been reported\textsuperscript{113}. This makes the increased CSF levels of Lynx1 of particular interest as it serves as a potential explanation of the increased activity of the cholinergic system. Thus, a possible mechanism of SCS presented in Paper V is that SCS induces dissociation of Lynx1 into CSF. Without usage of the stimulator, Lynx1 is normally expressed in the cell membrane and act as a brake for cholinergic activity. Upon electrical stimulation, Lynx1 is dissociated into the CSF, possibly by the increased levels of Sap3 that may hydrolyze the GPI anchor. As Lynx1 is then less able to carry out its modulating action on AChRs, the activity of the cholinergic system may be increased and less pain is experienced by the SCS user. The analgesic effect obtained by the user is known to last up to several hours after the stimulator has been turned off. The time it would take for Lynx1 to be transcribed and translated again into a functionally active protein in the cell membrane presents an attractive explanation to the long lasting analgesic effect. This is merely a hypothesis and further studies are needed to confirm this. The hypothesis does not explain how the actual pain mediated effect is produced nor does it explain the neuroanatomical basis. Regardless of the actual occurring mechanism, the higher concentration of Lynx1 in CSF upon SCS is an indication that this protein is linked to this mode of therapy. Thus, Paper V is the first study to report Lynx1 as a potential important molecular explanation of SCS-induced analgesia in humans.
8. Conclusion & future aspects

The endeavor in proteomics is to objectively analyze the entire proteome in an arbitrary sample. Analyzing thousands of proteins alone in a reliable and high throughput manner is an overwhelming technical challenge. Clinical proteomics take this a further step by also trying to translate the information into practical use. This adds an additional degree of complexity by also accounting for biological variation such as gender, age, ethnicity, exposure to environmental agents etc. Performing clinical neuroproteomics in the CNS as this doctoral thesis is focused on, combines these difficulties with the special challenges such as low protein amount, restricted sample availability and high dynamic range. Overall, this makes clinical neuroproteomics a very complicated discipline. Despite the challenges, the discovery of protein biomarkers holds great promise to facilitate diagnosis and treatment for CNS pathologies.

Currently, the most powerful strategy to screen for biomarkers in a global unbiased way is MS based proteomics. The methods used in Paper I-III may be used as tools for future protein biomarker discovery studies and their potential for protein enrichments have been demonstrated in combination with MS. The methods have been applied on biological material from the CNS but could also be applied to any biological matrix. In Paper I, the use of HLL was demonstrated. HLL seems to be suitable for a deep proteome profiling for samples with a high dynamic range. It should be noted that the acquired CSF volume in Paper I was much higher than would be assessable under more normal conditions. Thus, HLL seem to be more applicable in samples where higher protein content is available, i.e. plasma. A disadvantage with HLL is also the loss of quantitative information. Nevertheless, HLL may initially be used for a comprehensive qualitative analysis. The proteins of interest that have been identified from this first step may then be subjected for quantitative analysis in a follow-up step. This follow-up method may utilize a more targeted MS based approach, such as MRM that also have lower detection limits. The use of CPE for simultaneous protein enrichment of hydrophilic and hydrophobic proteins was first demonstrated in Paper II. The relative amount of membrane proteins identified was close to the reported genome expression level of 1/3 which makes this method especially suitable for characterization for this important group of proteins. It was then shown in Paper III that this technique is reproducible and allows for quantitative information. The method was also refined in terms of speed.
by using a different digestion protocol and combined with a better mass spectrometer. In summary, CPE was shown to yield an in-depth proteome analysis of brain tissue that may be used on large-scale sample sets.

The results in Paper IV demonstrate how the choice of protein quantification technique and mass spectrometer is important to obtain the best results from a proteomic experiment. Considerable differences were found between the methods and the strengths and weaknesses for each method are emphasized. Although MS based proteomics is a powerful technique, the limitations associated with it should not be neglected. Some of the artificially differentially expressed biomarkers were wrongly classified and some were not even identified at all. This demonstrates that the effort to characterize as many proteins as possible occurs at the cost of the analytical performance. Analyzing more biological and technical replicates will improve the results but MS based proteomics should be regarded as a screening method. Verification and validation of potential biomarkers by targeted MS methods or complimentary methods should therefore not only be preferable but should be required to ensure reliable results.

In Paper V, an assumption-free proteomic study was carried out in SCS-responsive humans to elucidate the underlying neurochemical mechanism of this mode of therapy. Seven proteins were found to have an altered concentration upon use of SCS in CSF. Lynx1 had the largest change in concentration and also had the most consistent change in all patients. Its link to the cholinergic system may be an important missing piece to the puzzle of how an analgesic effect is produced in the CNS by this mode of therapy. Further studies are required to validate these initial findings.

The end point in clinical proteomics is to establish an assay that may be used in clinical practice. Ultimately this assay will end up as a commercial product being sold to clinical routine laboratories. In contrast to the big effort put into this field, the rate of newly introduced protein assays on the market has declined over the past years to only about one per year\textsuperscript{119}. This demonstrates the long and difficult path from the discovery step to a novel clinical tool\textsuperscript{16}. Despite this long and difficult road, it is likely that many opportunities remain for applying new biomarkers, new technologies and refinements of existing methodology to future applications by protein analysis.
9. Summary in Swedish

9.1 Introduktion


Det centrala nervsystemet utgörs av hjärnan och ryggmärgen och omges av cerebrospinalvätska. Det centrala nervsystemet utgör en isolerad del av kroppen då denna del skämmas från resten av kroppen via blod-hjärnbarriären. Många folksjukdomar som t.ex. Alzheimer eller kronisk smärta är lokaliserade till detta område och en komplett förståelse eller behandling av dessa och många andra neurologiska sjukdomar saknas i dagsläget. Då det är mer sannolikt att hitta biomarkörer desto närmare man kommer till det drabbade sjukdomsområdet så utgör biologiska prover av cerebrospinalvätska och hjärnvävannad lovlige material för att framgångsrikt kunna finna sjukdomsrelaterade proteiner. Proteiner är dock en grupp av molekyler som är väldigt svåra att analysera. Vidare så försvåras analys av biologiska prover från det centrala nervsystemet av ett flertal faktorer då; (i) den totala koncentrationen av proteiner i dessa prover ofta är väldigt låg; (ii) har ett stort dynamiskt området, d.v.s. koncentrationsområdet som proteiner na förekommer i är väldigt brett; (iii) proverna är mycket komplexa med
tiotusentals olika proteiner; (iv) tillgängligheten av material från detta område är begränsat. Följdaktligen så behövs känsliga och specifika analysmetoder för denna uppgift. Denna doktorskavhandling är baserad på Artikel I-IV som berör framsteg för att kunna analysera proteiner från det centrala nerversystemet samt Artikel V som är en klinisk applikation.

9.2 Provupparbetning

Då ett biologiskt prov är mycket komplext som utöver proteiner även innehåller andra ämnen som lipider, salter kolhydrater m.m. så är det viktigt att först ha en effektiv upparbetningsmetod. Detta steg syftar till att extrahera och koncentrera önskade analyter (här proteiner) från övriga ämnen (matris).


Hydrofoba (ej vattenlösiga) proteiner utgör en typ av proteiner som är svåra att analysera då de flesta metoder är baserade på att proteiner är hydrofila (vattenlösiga). Hydrofoba proteiner är oftast membranproteiner som är förankrade till ytan av celler. Membranproteiner kan anses väldigt viktiga då den genetiska koden består så mycket som 1/3 av just membranproteiner samt att de flesta läkemedel inriktar sig just mot dessa proteiner. I Artikel II presenteras en metod för att analysera både hydrofila och hydrofoba proteiner samtidigt genom grumlingspunkts extraktion. Denna metod utnyttjar det faktum att den neutrala detergenten Triton X-114, kommer vid en viss temperatur (grumlingspunkt temperaturen) att bilda två distinkta faser. Genom att först inkubera hjärnväv med denna detergent så kommer de hydrofoba proteiner att lossna från cellmembranet och lösa sig i vattenfasen. Genom att sedan öka temperaturen till över grumlingspunkten kommer lösningen separera sig i två faser där den övre fasen innehåller hydrofila proteiner medan den nedre mer koncentrerade fasen består av hydrofoba proteiner som
har inneslutits inuti miceller som har bildats av detergenten. Olika protokoll för respektive typ av proteiner för vidare upparbetning kan sedan användas för att maximera antalet proteiner som kan analyseras. Denna metod demonstrerades framgångsrikt i grishjärna i Artikel II.


9.3 Masspektrometri

Masspektrometri (MS) är en av de mest kraftfulla detektionstekniker som existerar. Principen bygger på att joner i gasfas kan separeras och detekteras efter deras massa/laddnings förhållande. I princip vilken molekyl som helst kan mätas under förutsättning att analyten kan överföras till gasfas och bli elektriskt laddad. För att inte alla analyter ska introduceras samtidigt så är det viktigt att först ha en effektiv separation innan analyten når masspektrometer. MS är den viktigaste mättekniken inom proteomik och har använts konsekvent genom Artikel I-V. MS kan dels identifiera proteiner genom att utröna sekvensen av aminosyror som proteinet är uppbyggt av. Detta görs oftast genom att proteinet först klyvs av specifika enzymer ner till peptider som är mindre bitar av det intakta proteinet. Ifrån dessa mindre bitar kan sedan aminosyra sekvensen utläsas. Följaktligen så är det bara vissa delar av proteinet som verkligen detekteras. Vidare är MS oumbärlig för proteomik därför att det är den enda teknik idag som klarar av göra detta förutsättningsslööst och tillräckligt snabbt för tiousentals av peptider som kan introduceras till masspektrometer.

För att kunna finna kliniskt relevanta protein biomarkörer är det nödvändigt att jämföra den relativa koncentrationen mellan friska individer kontra sjuka patienter. Detta kan göras på flera olika sätt men görs antingen baserat på inmärkning eller inmärkningsfritt. Vid en inmärkning så introduceras olika stabila isotoper till respektive prover (sjuk/frisk). Efter inmärkning kan de båda proverna mixas och analyseras men masspektrometer kan fortfarande skilja peptider som ursprungligen kommer från olika individer då peptiderna skiljer i massa. I Artikel IV så undersöktes tre olika tekniker i kombination med två olika masspektrometrar om vilken som var bäst baserat på
flera olika urvalskriterier som är viktiga för att kunna identifiera biomarkö-
rer.

9.4 Klinisk applikation

Neuropatisk smärta definieras som smärta som uppstår som en konsekvens av en lesion eller sjukdom som påverkar det somatosensoriska nervsystemet. Farmaceutisk behandling är ofta inte tillräcklig och är bara effektiv för cirka hälften av patienter som har denna typ av nervskada. En alternativ behandlingsform är ryggmärgsstimulering där en elektrisk stimulator implanteras nära ryggmärgen. Elektriska impulser kan sedan utsändas av patienten genom en fjärrkontroll som ger en effektiv smärtlindrande känsla. Den moleky-
lära mekanismen för hur denna smärtlindring sker är fortfarande okänd.

I Artikel V gjordes en förutsättningslös undersökning om vilka proteiner som kan vara involverade i denna mekanism. Detta utfördes genom att 12 st frivilliga patienter som positivt svarar med denna behandlingsform avstod från att använda stimulat orn i 48 timmar. Efter denna period lämnade patien-
terna ett prov med cerebrospinalvätska. Därefter använde de stimulat orn som vanligt i 3 veckor och lämnade sedan ytterligare ett prov. Den metod för protein analys med MS som var bäst i föregående artikel användes i Artikel V. En panel av sju proteiner fanns vara statistiskt signifikanta för att korrele-
ra med användandet av stimulat orn. Proteinet Lynx1 uppvisade en särskilt stark korrelation då den genomsnittliga ökningen av koncentrationen var 59 % samt att den var förhöjd i 11 av 12 patienter vid användandet av stimula-
torn. Detta endogena protein uppvisar samma struktur och funktion som den aktiva delen i vissa ormgift. Lynx1 sitter förankrat till cellmembran via ett kemiskt ankare där den utövar en modulerande funktion på acetylkolinrecep-
torer. En förhöjning av acetylkolin koncentrationen har tidigare rapporterats vid denna behandlingsform. En möjlig teori är att när stimulat orn används så lossnar ankaret från proteinet och dissocierar därmed från cellmembranet ut i cerebrospinalvätskan. När Lynx1 inte längre finns i närheten av acetylkolinreceptorn där den utövar sin funktion så kommer aktiviteten av denna recep-
tor att förändras. Detta är bara en hypotes och mycket arbete återstår för att bekräfta denna teori. Artikel V är den första studien som rapporterar Lynx1 som en möjlig molekylär förklaring till hur smärtlindring via ryggmärgssti-
mulering signaleras i människor med neuropatisk smärta.
10. Acknowledgement

This work would not have been possible without the aid from colleagues and support from friends and family. I would like to express my deep gratitude to some selected individuals for contributing to this work.

To my main supervisor Magnus Wetterhall, for your devoted (occasionally fanatic!) supervision and substantial contribution in all aspects of this doctoral thesis. But also for setting the standard for how a true friend should be.

To my co-supervisor Jonas Bergquist for believing and giving me the opportunity to initiate my science studies.

David Malmström, you are the best office roommate one can ask for and a good friend and colleague. Thank you for your all computer related assistance, help with graphics and good discussions (not always work related!).

Thank you Andreas Dahlin for your professional help with the graphics.

Titti Ekegren, Jörg Hanrieder, Ganna Shevchenko & Konstantin Artemenko. Each and one of you have contributed by your specific area of expertise.

Anne-Li Lind, for making labwork even more fun and a fruitful collaboration. Kim Kultima and Rolf Danielsson you both have contributed with indispensable help with statistics and being a source of inspiration.

My family to whom this thesis is dedicated to! Lukas, my dear son, this thesis has been important for me to write for a period of my life, but you will always be the most important in my entire life! My mother Eva-Marie, for your never ending support and kindness. My brother Andreas, for probably initiating my career by convincing me to aim for the highest degree in chemistry already in the 7:th grade.

Last but definitely not least, I would like to extend a big gratitude to all the wonderful people working at the Department of Analytical Chemistry in Uppsala. You all make it a pleasure to go there every morning.
11. References


Shevchenko, G., et al., Cloud-point extraction and delipidation of porcine brain proteins in combination with bottom-up mass spectrometry approaches for proteome analysis. *Journal of Proteome Research* 9, (8), 3903-3911.


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology.