Karolinska Institutet

Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

**Doctoral Thesis in Medical Science**

# Proteomics of Neurodegenerative Diseases Using Novel Online Isoelectric Point Separation

Mohammad Pirmoradian

Stockholm 2015

© Stockholm 2015, Mohammad Pirmoradian Najafabadi
Printed By Universitetsservise US-AB
ISBN 978-91-7676-159-5

All articles are reproduced with permission from the publishers.

# Proteomics of Neurodegenerative Diseases Using Novel Online Isoelectric Point Separation

Thesis for Doctoral Degree (Ph.D.)

*Respondent:*
**Mohammad Pirmoradian**
Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Chemistry I

*Principal Supervisor:*
**Prof. Roman A. Zubarev**
Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Chemistry I

*Co-supervisor(s):*
**Prof. Dag Årsland**
Karolinska Institutet, Department of Neurobiology, Division of Translational Geriatric Neuropsychiatry

**Prof. Thomas Helleday**
Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Translational Medicine, Science for Life Laboratory

*Opponent:*
**Prof. Neil Kelleher**
Northwestern University, Departments of Chemistry and Molecular Biosciences, and the Proteomics Center of Excellence

*Examination Board:*
**Prof. Pär Nordlund**
Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Biophysics

**Assoc. Prof. Janne Lehtiö**
Karolinska Institutet, Department of Oncology-Pathology, Division of Clinical Proteomics Mass Spectrometry, Science for Life Laboratory

**Assoc. Prof. Jochen M. Schwenk**
KTH – Royal Institute of Technology, School of Biotechnology, Division of Proteomics and Nanobiotechnology, Science for Life Laboratory

*Chair of the session:*
**Assoc. Prof. Josefin Skogsberg**
Karolinska Institutet, Department of Medical Biochemistry and Biophysics, Division of Vascular Biology

***To my family***

# Abstract

The work presented in this thesis describes an instrument, developed for separation of proteins and peptides based on corresponding isoelectric point (pI) values, to empower mass spectrometry-based proteomics analysis. The main objectives are the instrument development and optimization, as well as clinical applications in biomarker discovery, particularly for neurodegenerative disorders. The thesis is based on five scientific papers that focus on three main stages; (i) development and optimization of the device for separation of proteins and peptides by pI that are well integrated with tandem mass spectrometry (ii) optimization of the device for intact blood plasma protein fractionation for application in biomarker discovery, and (iii) biomarker discovery in blood plasma for early diagnosis of Alzheimer disease.

Within the first part of the work, a novel multiple-junction capillary isoelectric focusing fractionator (MJ-CIEF) is developed (**Paper I**). Subsequently, the resolving power and reproducibility of fractionation are improved, and in addition, a novel algorithm is developed to calculate the identified peptides' pI (**Paper II**). Moreover, to achieve the aim of deep proteomics, a multi-parameter optimization of the LC-MS/MS pipeline is performed (**Paper III**).

In the second part, an online desalinator is developed and coupled to the device, for direct buffer-exchange and isoelectric separation of intact human blood plasma/serum proteins. The developed pipeline achieves the increased depth of the proteome analysis and provides additional information on the pI of identified proteins, as another dimension of information in biomarker discovery by proteomics (**Paper IV**).

The last part of the thesis is focused on the application of the developed method in biomarker discovery for early diagnosis of Alzheimer disease. A panel of new potential biomarkers is introduced based on the abundance changes, as well as shifts in the pI values. By means of the pI information, the protein concentration in a narrow pH range around 7.4 reveals increased levels in patients with progressive Alzheimer disease compared to stable ones. Proteome analysis of this particular pI region also suggests several potential proteins as biomarkers for early diagnosis of the disease (**Paper V**).

Taken together, this thesis demonstrates emerging applications of peptides and proteins fractionation by pI in deep proteomics. The development of MJ-CIEF facilitates online separation of peptides and proteins from small amounts of samples in a fast format, automatable, cost-effective and compatible with mass spectrometry analysis. Further biomarker discovery in the narrow range of pH around 7.4 of blood proteome is suggested for early diagnosis of neurodegenerative diseases.

# Sammanfattning

Denna avhandling beskriver metodutvecklingen av ett instrument som utvecklats för att separera proteiner och peptider via fokusering av molekylernas Isoelektriska punkt värden (Ip), följt av masspektrometri-baserad proteomik analys. Förutom instrumentutveckling och optimering var syftet att testa och applicera metoden på kliniska prover för att identifiera nya biomarkörer för neurodegenerativa sjukdomar. Avhandlingen är baserad på fem vetenskapliga artiklar som fokuserar på tre huvudkriterier; (i) utveckling och optimering av ett instrument som separerar proteiner och peptider genom deras pI-värden för vidare analys via en direktkopplad tandem-masspektrometer, (ii) optimering av instrumentet för intakt blodplasma protein fraktionering, och (iii) applicering av instrumentet för att upptäcka biomarkörer i blodplasma för tidig diagnos av Alzheimers sjukdom.

Den första delen av avhandlingen, beskriver utvecklingen av det nya isoelektrisk fokusering instrumentet (MJ-CIEF) (**Artikel I**), samt hur resolutionen och reproducerbarheten av fraktioneringen har förbättras. En ny algoritm har utvecklats för att beräkna de identifierade peptidernas lp värden (**Artikel II**). Dessutom, en multi-parameter optimering av en LC-MS/MS pipeline genomfördes med syftet att nå en större del av det totala proteomet (**Artikel III**).

I del två av avhandligen beskrivs metodutvecklingen av ett online avsaltnings system som kopplas till instrumentet för direkt buffert-utbyte och isoelektrisk separation av intakta humana blodplasma proteiner. Tillämpningen av den utvecklade pipelinen innebär att proteomet kan analyseras mer detaljerat. Dessutom ger metoden kompletterande information i form av proteinernas lp värden, vilket ger ytterligare en dimension av information för respektive protein/peptid som kan vara av vikt speciellt för att upptäcka nya biomarkörer (**Artikel IV**).

Den tredje och sista delen av avhandlingen fokuserar på tillämpning av den utvecklade metoden för att upptäcka biomarkörer för tidig diagnos av Alzheimers sjukdom. En panel av nya potentiella biomarkörer identifierades, som var baserad på storleksförändringar, liksom förändringar i lp av proteiner. Specifikt i ett smalt pH-område runt 7.4 upptäcktes en specifik proteinprofil i patienterna med progressiv Alzheimers sjukdom. Analys av proteomet för denna region identifierade flera proteiner som potentiellt kan användas som biomarkörer för tidig diagnos av sjukdomen (**Artikel V**).

Sammanfattningsvis, beskriver denna avhandling ett nytt tillvägagångsätt för att karaktärisera och profilera peptider och proteiner för djup proteomik analys. Utvecklingen av MJ-CIEF förenklar online separering av peptider och proteiner från små mängder av prover på ett snabbt, automatiserbart och

kostnadseffektiv sätt som är kompatibelt med masspektrometrianalys. Dessutom, resultaten från vår pilot studie indikerar att metoden kan användas för att tidigt diagnosera neurodegenerativa sjukdomar från specifika proteiner inom ett snävt pH-område runt 7.4 i blodplasma proteomet.

# List of Scientific Publications

**PAPER I**

Chingin K., Astorga-Wells J., Pirmoradian M., Lavold T., Zubarev R. A. (2012) Separation of Polypeptides by Isoelectric Point Focusing in Electrospray-Friendly Solution Using a Multiple-Junction Capillary Fractionator. Analytical Chemistry 84: 6856-6862.

**PAPER II**

Pirmoradian M.*, Zhang B.*, Chingin K., Astorga-Wells J., Zubarev R. A. (2014) Membrane-Assisted Isoelectric Focusing Device As a Micropreparative Fractionator for Two-Dimensional Shotgun Proteomics Analytical Chemistry 86 (12), 5728-5732.

**PAPER III**

Pirmoradian M., Budamgunta H., Chingin K., Zhang B., Astorga-Wells J., Zubarev R. A. (2013) Rapid and Deep Human Proteome Analysis by Single-Dimension Shotgun Proteomics. Mol Cell Proteomics 12: 3330-3338.

**PAPER IV**

Pirmoradian M., Astorga-Wells J., Zubarev R. A. (2015) Multijunction Capillary Isoelectric Focusing Device Combined with Online Membrane-Assisted Buffer Exchanger Enables Isoelectric Point Fractionation of Intact Human Plasma Proteins for Biomarker Discovery. Analytical Chemistry, DOI: 10.1021/acs.analchem.5b03344.

**PAPER V**

Pirmoradian M., Årsland D., Zubarev R. A. Blood Proteins with Isoelectric Point close to 7.4 as Alzheimer Disease Biomarker. Manuscript.

* equal contributions

# Scientific Publications (not included in this thesis)

Zhang B., Pirmoradian M., Chernobrovkin A., Zubarev R. A. (2014) DeMix Workflow for Efficient Identification of Cofragmented Peptides in High Resolution Data-dependent Tandem Mass Spectrometry. Molecular & Cellular Proteomics 13 (11), 3211-3223.

Saei AA., Chernobrovkin A., Sabatier P., Pirmoradian M., Zubarev R. A. Proteomic characterization of target and mechanism of action in response to small-molecule anticancer drugs. Manuscript.

Abbasi M., Pirmoradian M., Yang H., Zubarev R. A. Asparagine deamidation and isoaspartyl residues can induce reactive oxygen species. Manuscript.

# Table of Contents

**Chapter 1 Background** .......... **1**
Proteomics in Medical Science .......... 3
Mass Spectrometry in Protein Analysis .......... 5
Separation Techniques .......... 7
*Electrophoresis* .......... *10*
*Capillary Electrophoresis* .......... *11*
*Isoelectric Focusing* .......... *12*
*Capillary Isoelectric Focusing* .......... *14*
Deep Proteomics .......... 16
Blood Plasma Proteome .......... 20
Alzheimer Disease .......... 23

**Chapter 2 Present Investigations** .......... **27**
Main Objectives .......... 28
*Specific Aims* .......... *28*
Methodology .......... 29
*Blood Samples* .......... *29*
*Ethical Permits* .......... *29*
*The MJ-CIEF Device* .......... *30*
*Fractionation Workflow* .......... *31*
*Desalinator* .......... *32*
*Protein Digestion* .......... *33*
*RPLC-MS/MS* .......... *33*
*Data analysis* .......... *34*
Results and discussion .......... 36
*Paper I- Developing an online multiple-junction capillary isoelectric focusing fractionator (OMJ-CIEF) device* .......... *38*
*Paper II- Optimizing the MJ-CIEF device as a micro-preparative fractionator* .......... *40*
*Paper III- Rapid and deep single-dimension shotgun proteomics* .......... *43*
*Paper IV- Online desalination and isoelectric focusing of plasma proteome* .......... *46*
*Paper V- Narrow pH range of 7.4 in biomarker discovery of neurodegenerative disorders* .......... 48

**Chapter 3 Conclusion and Future Perspectives** .......... **51**
Conclusion and Future Perspectives .......... 52

**Acknowledgment** .......... **55**
Acknowledgment .......... 56

**Bibliography** .......... **61**
Bibliography .......... 62

**Appendix** .......... **71**

## List of Abbreviations

| | |
|---|---|
| A2M | Alpha-2-macroglobulin |
| AD | Alzheimer disease |
| APP | Amyloid precursor protein |
| Aβ | Amyloid beta |
| BSA | Bovine serum albumin |
| CA | Carrier ampholyte |
| CE | Capillary electrophoresis |
| CID | Collision-induced dissociation |
| CIEF | Capillary isoelectric focusing |
| CSF | Cerebrospinal fluid |
| CV | Coefficient of variation |
| DDA | Data-dependent acquisition |
| DEW | Dynamic exclusion mass windows |
| DIA | Data-independent acquisition |
| DNA | Deoxyribonucleic acid |
| DTT | Dithiothreitol |
| EOF | Electroosmotic flow |
| ESI | Electrospray ionization |
| FA | Formic acid |
| FDR | False discovery rate |
| FT-ICR | Fourier transform ion cyclotron resonance |
| GO | Gene ontology |
| Hb | Hemoglobin |
| HCD | Higher-energy C-trap dissociation |
| HPLC | High-performance liquid chromatography |
| IAA | Iodoacetamide |
| ICR | Ion cyclotron resonance |
| ID | Inner diameter |
| IEF | Isoelectric focusing |
| IPG-IEF | Immobilized pH gradient isoelectric focusing |
| isoAsp | Isoaspartate |
| LC-MS/MS | Liquid chromatography-tandem mass spectrometry |
| m/z | Mass-to-charge ratio |

| | |
|---|---|
| MALDI | Matrix-assisted laser desorption ionization |
| MMSE | Mini–mental state examination |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| (O)MJ-CIEF | (Online) multiple-junction capillary isoelectric focusing |
| OD | Outside diameter |
| PMCI | (Progressive) mild cognitive impairment |
| PD | Parkinson disease |
| PEEK | Polyether ether ketone |
| pI | Isoelectric point |
| PSEN | Presenilin |
| PTM | Post-translational modification |
| RNA | Ribonucleic acid |
| RPLC | Reversed-phase liquid chromatography |
| SDC | Sodium desoxycholate |
| SDS-PAGE | Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis |
| SPE | Solid phase extraction |
| SUMO | Small ubiquitin-like modifier |
| TFA | Trifluoroacetic acid |
| TOF | Time-of-flight |
| tR | Retention time (in chromatography) |
| UV | Ultraviolet |

# Chapter 1
# Background

# Proteomics in Medical Science

Medicine is the science of preserving and restoring health. Medical sciences are dedicated to maintaining the well-being of humankind by emerging skills in prevention, diagnosis, alleviation and treatment of diseases. The cornerstone of medicine has been based on detecting the indications of any deviation from the state of health. As many other basic sciences, advances in medicine are founded on careful examinations and observations, including measurements of all possible markers of the diseased state. Conventional medicine was historically based on evident physical symptoms of disease incident, and studies often directly involved human subjects. Recent advances in science, technology and informatics, as well as the access to the disease models, dramatically accelerated the developments in medical science and clinical practice and brought the depth of observations to the molecular level.

The markers of health or diseased states are now referred to as biomarkers and have a categorized definition. A biomarker is an analyte or a condition of the body that is measurable and is shown to be closely associated with the state of health. Biomarkers can be indicative of a present state or predictive for a future outcome.[1] Nowadays, biomarkers not only aim for early diagnosis but are also applied to monitor the prognosis of diseases or the success of the response to a particular treatment. Advances in cell and molecular sciences provided countless opportunities and enhanced the amount of knowledge in modern medicine. To date, biomarkers at molecular level such as genes and DNA, different types of RNAs, proteins and metabolites, enable early diagnosis with improved accuracy and can predict the occurrence of diseases, even before an individual is born.

Among many biomolecules, proteins are the critical functional actors in cellular processes and are involved in numerous disease mechanisms. Hence, several proteins represent as a marker of disease onset or progression and are among the main targets of drugs.[2] The word protein was first used by the analytical chemists Jöns Jacob Berzelius and Gerardus Mulder, from the Greek root proteios meaning primary, emphasizing the fundamental importance of these molecules.[3]

Proteins are made up of 20 standard monomers called amino acids that offer enormous variability in the polypeptide chain of proteins. This variation is not limited by the number of corresponding genes. Originating from the same gene, several possible isoforms can be translated into the final amino acid sequences, when the corresponding RNA gets spliced into diverse isoforms. Additionally, proteins can fold into different 3-D structures that can bring distant amino acids together to form unique structural and functional domains. While the primary structure of proteins is the order of amino acids, physical and chemical properties of amino acids dictate the secondary structures such as turns, sheets and helixes. These local structures fold the tertiary structure of proteins, which alone or as a combination of several complementary subunits (quaternary structure) shape the entire 3-D structure of the protein.[4] Beyond that, the amino acid residues of a protein can be chemically modified, which is known as post-translational modifications (PTMs). More than several hundreds of such modifications are known to alter proteins in human such as phosphorylation, glycosylation, deamidation, acetylation, hydroxylation, methylation and ubiquitylation. PTMs affect the ultimate protein conformation, function or localisation.[5] Considering all possible variations in the final protein products of a gene, the term proteoform was introduced to designate each of the molecular forms.[6]

Proteins play different roles and perform a wide range of tasks based on their structural and physicochemical properties. Therefore, abnormal deviation in protein properties can manifest diseases. Accordingly, analysis of proteins and their associated properties are of significant interest in biological and medical research.[7] A protein state can be represented by its measurable chemical and physical properties such as size and molecular weight, affinity and interaction with other proteins or chemicals, lack or gain of a chemical modification, enzymatic activity, isoelectric point, melting point, and

copy numbers or abundance of that protein. On the other hand, proteins orchestrate together in the cellular and organs' actions. Therefore, the study of the entire repertoire of expressed proteins, i.e. the proteome, is also significantly important.[8]

A proteomics study is defined as the examination of the complete set of proteins at a specific state, time or location. This global approach transforms the biomedical research into an investigation of the entire cellular system. The attempts towards that objectives go back to 1970s with the introduction of two-dimensional (2-D) polyacrylamide gel electrophoresis.[9] However, actual achievements to approach the goal had to wait for the development of mass spectrometry (MS), which enabled unbiased protein detections with theoretical specificity and sensitivity at the level of single molecules.[10-12]

## Mass Spectrometry in Protein Analysis

Mass spectrometry (formerly known as mass spectrography) was introduced by Joseph John Thomson in 1897 in an experiment in which he studied the movement of negatively charged cathode ray particles (electrons), as well as other charged particles, in electromagnetic fields.[13] He demonstrated that the mass-to-charge ratios (m/z) of particles define their movement in electric and magnetic fields. Rapidly, the value of mass spectrometry application was recognized in different fields of science, particularly the analytical sciences. Vast amount of advancements have been made in the field of mass spectrometry, and variety of techniques have been developed, which is still actively ongoing.

In principle, each mass spectrometer has three major units known as the ion source, mass analyzer, and the ion detector. It also has complement units of inlet system for the ion source, vacuum system, electronics control board and data processor. The ion source unit ionizes and transfers the analytes into the gaseous phase. Methods for ionizations that were in use decades ago were strong activation processes such as electron ionization, photoionization, atom bombardment, as well as less harsh chemical ionization. These methods were not well suited for biomolecules such as proteins that are fragile and could decay as a result of ionization. This issue was resolved by the development of more gentle, "soft" ionization methods called electrospray ionization (ESI) and

matrix-assisted laser desorption ionization (MALDI), by John B Fenn and Koichi Tanaka who shared the 2002 Nobel Prize in chemistry.

By the action of electric potential and pressure differences, the ionized molecules, formed in the ionizer, will flow into the mass analyzer unit. In the mass analyzer, the ions' motion will be influenced by electric and/or magnetic fields that are associated with their characteristic mass-to-charge ratios (m/z). Several mass analyzers have been developed with differences in ion separation and analysis. Time-of-flight (TOF) and ion cyclotron resonance (ICR) analyzers were first introduced in 1946 and 1948 respectively.[14,15] ICR requires strong magnets and thus expensive. In contrast, the quadrupole filter and ion trap, developed by Wolfgang Paul (therefore also known as Paul trap), are more available. [16,17] Other analyzers include magnetic sector and Fourier transform (FT) generic types, such as Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap.[18]

The main parameters of the mass analyzer are mass accuracy, m/z range, dynamic range, sensitivity, resolution, speed and the ability to perform fragmentation. Often two analyzers are combined, in order to improve the analysis power, which is frequently used in proteomic applications. In this case, the first mass analyzer and fragmentation is followed by the second analyzer (tandem mass spectrometry, MS/MS). In tandem MS, a particular peptide ion, known as precursor ion, is isolated and fragmented, typically by collisions with an inert gas such as nitrogen or helium, resulting in product ions. The mass spectrum of product ions is called tandem MS, also abbreviated as MS/MS or MS2. This methodology (MS/MS) is a key technique for peptide identification in MS-based proteomics.[19] Collision-induced dissociation (CID)[20] and the higher-energy collisional dissociation (HCD)[21] are among widely used techniques for fragmentation. Moreover, development of the complementary fragmentation techniques electron capture dissociation (ECD)[22] and electron transfer dissociation (ETD)[23] facilitated the analysis of PTMs such as glycosylation and phosphorylation.

Ions, separated in the mass analyzers by their m/z, produce electrical signals. In the detector unit, these signals are detected and amplified in order to reach a better sensitivity. The data is presented as mass spectra, with the x-axis being m/z, and the y-axis representing relative signal intensities.

As previously noted, proteomics studies deal with the analysis of highly complex samples, where solely the differences in protein abundances can exceed 10 orders of magnitude.[24] Considering the current limits of mass spectrometers instrumentation in terms of ion capacity and dynamic range (3-4 orders of magnitude), such complexity needs to be reduced prior to MS analysis, in order to reach a deeper characterization of proteomic samples. In that regard, the application of separation and enrichment techniques to provide subsets of protein mixtures in separate fractions with reduced complexity plays a key role in sample preparation step for MS-based proteomics studies.

## Separation Techniques

As mentioned above, proteins are biomolecules with very diverse biochemical characteristics, including the sequence of amino acids, size, PTM, 3-D structure and solubility. On one hand, this makes the proteome highly complex and challenging to analyze. However, on the other hand, these differences can be used to separate proteins into subsets in order to simplify the subsequent analysis.

Separation can be based on biological or physicochemical properties. Biological properties are usually used for preparative fractionations aiming at the enrichment of target proteins. The common methods include sub-cellular fractionation[25], separations based on interaction to other proteins[26,27] or nucleic acids,[28,29] PTMs like phosphorylation[30-33], acetylation[34], glycosylation[35] and SUMOlyation[36]. Separations based on physicochemical properties are closely related to the sequence and properties of the proteins' amino acids. Often, separation based on one single property is not sufficient to cover the vast complexity of proteome, and thus two methods will be combined to perform a two-dimensional separation.

The most commonly used separation method that is being applied in MS-based proteomics is reversed-phase high-performance liquid chromatography (HPLC). Amino acids with a hydrophobic side chain such as tryptophan (W), phenylalanine (F), leucine (L), isoleucine (I), valine (V), and tyrosine (Y) have the biggest effect on hydrophobicity of the proteins and peptides.[37] The most common stationary phase in this separation is octadecyl carbon chain

(C18)-bonded and C4-bonded silica for peptides and proteins respectively.[38] In LC-MS/MS, the chromatography column is coupled to the ESI source of the mass spectrometer.[39] The acidic environment of separation in HPLC is advantageous for ionization using positive ESI mode.

Another commonly practiced technique for sample pre-fractionation prior to MS analysis is SDS-PAGE. The loaded protein samples on the gel are separated by size. After cutting the gel, fractions are digested in the gel using protease enzymes such as trypsin.[40,41] Digested peptides are extracted, purified and are analyzed by LC-MS/MS. Proteins or peptides can also be separated using ion exchange chromatography that can be combined orthogonally to reversed-phase liquid chromatography (RPLC).[42]

To achieve the highest resolving power for proteomics applications,[43] proteins or peptides in a complex sample can be separated based on their isoelectric point (pI) value. The isoelectric point of a protein or peptide is the equivalent to the pH of a buffer, in which the net surface charge of the molecule is zero. The pI of proteins therefore is mostly dependent on the amino acids with charged side chains such as glutamic acid (E), aspartic acid (D), histidine (H), lysine (K) and arginine (R), as well as protein terminal groups (NH2 at the N-terminal and COOH at the C-terminal ends). Classical theoretical formula for pI calculation is based on Henderson–Hasselbalch equation,[44,45] in which the pKa of the acid dissociation constant of charged amino acids in the protein determines the pH in which the total charge ($C_{total}$) equals zero:

for negatively charged residues: $C_N = \sum_{i=1}^{N} \frac{-1}{1+10^{pK_{N_i}-pH}}$

for positively charged residues: $C_P = \sum_{i=1}^{P} \frac{1}{1+10^{pH-pK_{P_i}}}$

$$C_{total} = C_N + C_P$$

Moreover, the influence of other amino acids, sequence and structural properties, as well as chemical modifications of the protein and peptide are considered in more advanced pI calculation formulas.

In practice, proteins and peptides are separated based on their pI by isoelectric focusing (IEF).[46] IEF is compatible with numerous other separation methods like chromatography and gel electrophoresis; therefore, it can be readily combined for 2-D separations. A well-known example is 2-D gel electrophoresis, in which separated proteins by IEF are subjected to size separation in the second dimension.

An additional advantage of using IEF in the analysis of proteins is the separation of proteins in individual fractions based on their pI value. The isoelectric point of a protein is an important and informative parameter that reflects the intrinsic properties of the proteins as well as its PTMs. Moreover, the overall distribution of proteins in a complex protein mixture (such as human blood) can be investigated over the theoretical pI values. One intriguing instance of such analysis regards the pH range from 7.30 to 7.50 in human blood plasma. An in-silico analysis of proteins' pI values suggests that this range is a forbidden region for almost all proteins in the human blood proteome. Very few normal protein isoforms appear to have a pI in this range. This can be interpreted as the increased risk of neutralization and precipitation at the pH of blood (7.4) that causes normal proteins to avoid this pI-value (Figure 1).

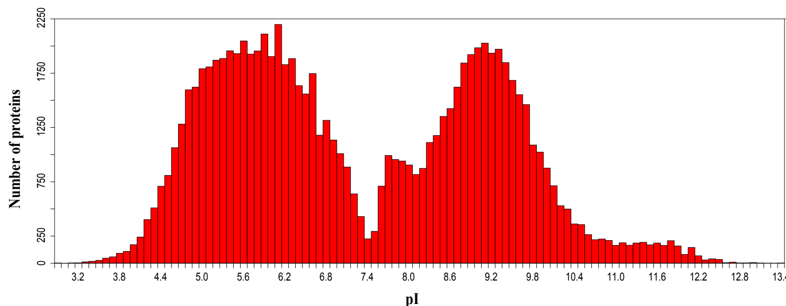


**Figure 1**. **pI distribution of human proteome**. The amino acid sequences of all protein isoforms from International Protein Index (human version 3.87) were retrieved. The theoretical isoelectric point for each sequence was calculated in R by means of "seqinR" package. Then the pI distribution was plotted in this histogram. (Similar distribution was previously reported for the human proteome[47])

Considering the importance and vast application of separation methods, particularly electrophoresis-based techniques, in MS-based proteomics studies, the history and fundamentals behind the development of these techniques are discussed in more detail in the following sections.

## Electrophoresis

Arne Tiselius, a Swedish biochemist and Nobel laureate, obtained his Ph.D. in Svedberg Lab at Uppsala University, Sweden. Theodor Svedberg, who also won the 1926 Nobel Prize in chemistry, had realized that colloids could be separated by migration through electric fields.[48] This process was called electrophoresis. Tiselius continued working on electrophoresis during his Ph.D. and nailed his doctoral thesis in 1930, entitled "The Moving Boundary Method of Studying the Electrophoresis of Proteins".[49] Tiselius improved the U-tube apparatus for electrophoresis that allowed him to separate serum proteins. Since the separations of some artificial mixtures did not produce distinct bands, so the utility of such methods at the time was not clear. Years later, Tiselius redesigned the electrophoretic U-tube and demonstrated the improved resolution on a sample of serum. For that, he dialyzed the serum using a buffer solution and after 2 hours of running the separation, the serum had resolved into 4 bands referred to as albumin, α-, β-, and γ-globulins.[50] Tiselius submitted the corresponding paper to a biochemical journal, which was rejected as it was too "physical".[51] However, the valued paper ended up in the Transactions of the Faraday Society[50] and later in 1948 along other astonishing works gained him the Nobel Prize in chemistry. This cornerstone of a new analytical tool in biochemistry was developed further in Tiselius Lab by contribution of pioneers, mostly of his Ph.D. students.[52]

The main concept of electrophoresis is based on separation of charged molecules in an electric field. Depending on the pH of the environment, proteins and peptides can have different net charges on their surface. In the presence of an electric field, charged molecules feel the force and each one moves with a velocity proportional to the mass-to-charge ratio (m/z) of that molecule. Having different masses and charges, molecules have different m/z and, therefore, different velocities to move in a constant electric field. As a result, differences in the velocities can separate amphoteric molecules such as proteins during an electrophoretic run.[53]

Several different formats of electrophoresis have been developed, in which the principle of separation is similar, yet the separation channel is diverse. Instances of such methods include capillary electrophoresis (CE), isoelectric focusing (IEF), in-gel and off-gel electrophoresis and open channel electrophoresis. These configurations are intended to address different issues in

the application of the technique for different separation purposes, such as decreasing the sample amount and reducing the heat joule production in the capillary electrophoresis, separation of amphoteric molecules based on pI value in a pH gradient in IEF, separation of proteins based on size in gel electrophoresis, and facilitating sample collection from gels in off-gel formats. [54]

## Capillary Electrophoresis

Tiselius and his group's works on electrophoresis were the cornerstone in understanding the details of this method, which shed light on the direction of its development during the following years, that was facilitated by the emerging advanced materials.[55] The diameter of the separation column decreased to the scale of micrometer, and the fused silica capillary columns became the dominant material of choice for capillary columns in CE.[56] Fused silica has free hydroxyl groups on the surface that are ionized during electrophoresis experiment and generate a bulk flow of solution in the capillary. This phenomenon, known as electroosmotic flow (EOF), changes the laminar flow profile to a plug profile.[53]

Among electrophoretic separation techniques, CE has several advantages including the lower Joule heating effect. This is due to the large surface to volume ratio in the CE column that helps to dispatch the heat. Another advantage is the low conductivity of the column that allows for the use of relatively high electric field strength (100–500 V/cm) without the concern for heat generation. The possibility of applying higher electric field results in better resolution and efficiency in shorter running times. Furthermore, CE (in common with chromatography techniques) has the benefits of minimal sample size, ease of sample preparation, cost-effectiveness as well as adaptability to method development and automation.[53]

CE is an excellent tool for the analysis of DNA (mainly due to its native negative charge that is repelled by the negatively charged fused silica surface), [57] pharmaceutical products (chiral analysis)[58] and metabolites. However, separation of protein samples with CE is still challenging. This is partially due to the amphoteric characteristic of proteins. Moreover, the protein molecules interact with hydroxyl groups of silanols in fused silica that results in the stickiness of the sample to the fused silica wall of the capillary. During the

years, many coatings for the column have been introduced. The available coatings work by three main approaches of quenching the interactions (by ionic strength or competing ions), micro-ion adsorption to the capillary walls, or covalent modification of silanols with polymers. [59] However, none of the above is the perfect solution and the separation of proteins, particularly in a complex context, via CE is still a challenge.

## Isoelectric Focusing

Isoelectric focusing is an electrophoresis technique but in a pH gradient instead of a constant pH along the separation field. The separation then is based on the isoelectric point of molecules. In other words, the analyte molecules such as proteins move towards the electric field according to their net charge until they reach the zone, in which the pH of the environment is equal to their pI and, therefore, the net of charge is zero. Unlike electrophoresis, peaks are self-sharpened during IEF separation due to the increase in charge by moving away from the focusing position and consequently being pushed back by the electric field (Figure 2).

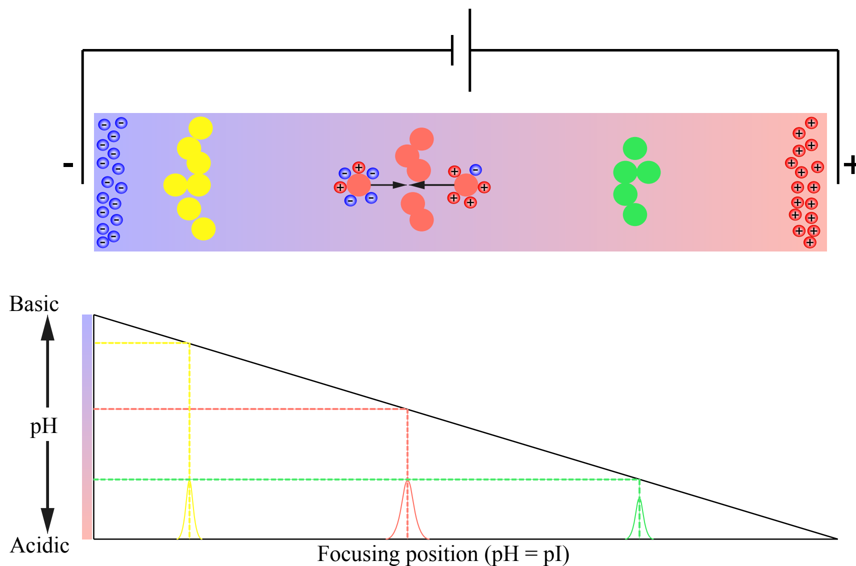


**Figure 2**. **Isoelectric Focusing**. Three proteins are schematically shown in yellow, orange and green circles; each is focused in a position where the local pH is equal to its pI. Moving away from that position makes the protein be charged and, therefore, be forced back to the focusing position by the electric field.

As explained in the concept of IEF, maintaining a stable pH gradient is the essential component of a successful IEF. The roots of the first attempts on performing IEF also goes back to the Tiselius Lab where Harry Rilbe (formerly Harry Svensson) attempted to make the gradient by trying different buffers during his Ph.D. work. The problem was that the ionic constituents of buffers were completely migrating to the opposite charge electrodes. This caused a non-ionic pure water region in the middle with very low electrical conductivity, resulting in the liquid to almost start to boil. He tried to solve this issue by continuous supplying of fresh buffers into the giant apparatus during the separation, and still the method did not make any stable pH gradient. [60] Years later, he attempted to create a pH gradient by mixing of a series of buffers that he called "carrier ampholytes", all of which had to be amphoteric with a small value of $\Delta$pK in order to ensure the buffering power.[61] With such a mixture, the electric field could focus them to these isoelectric points. Diffusion would broaden them and slightly make them penetrate into the neighboring zones, where they ensure electric conductivity and maintain the electrical current along the separation column. [62,63]

During his work at Karolinska Institutet, Svensson together with his medical student Olof Vesterberg, designed experiments for proper synthesis of carrier ampholytes (CA). Three years later in 1964, Vesterberg synthesized a mixture of random oligoamines (from tetra- to hexa-amino groups reacted with acrylic acid).[64] This invention resulted in the initiation of a broad application of IEF in separation science. After a decade improvements of the technique, in 1975, applying the IEF as the first dimension of 2-D gel electrophoresis, the method was permanently established in the realm of biochemical sciences, that even today is still one of the most popular techniques practiced in proteome analysis.[9] Nowadays, there are several commercial carrier ampholytes available, covering narrow and broad ranges of pH scale. Most important parameters for a successful IEF separation performance are providing a stable and linear pH gradient as well as electric field along the separation zone.

In 1982, the concept of immobilized pH gradients was introduced, in which ampholyte compounds were fixed in a gel (immobilized pH gradient - IPG-IEF). This could solve some of the problems associated with carrier ampholytes in solution, such as uneven buffering capacity and conductivity, irreproducibility of CA synthesis as well as cathodic drift during the focusing.[65]

Today, there are well-developed techniques for high-resolution pI separation of proteins and peptides using IPG-IEF separation, which are applied in proteomics studies.[43] Nevertheless, the gel-based methods are still time consuming and expensive compared to in-solution IEF, are less automated and often require considerably larger initial sample amount for reaching the same recovery of separated analytes.

## Capillary Isoelectric Focusing

In contrast to capillary electrophoresis with a single-step separation procedure, capillary isoelectric focusing (CIEF) is performed in a two-step process. First, the focused proteins will remain steady at the position, in which their pI equals the pH zone, which is called the focusing step. Unlike CE, the electroosmotic flow should be avoided, and for that the capillary in CIEF is often coated with a hydrophilic coating such as polyacrylamide. The second step of CIEF procedure is the detection or collection of fractions, in which the separated molecules should be mobilized towards the detector (often positioned at the cathodic end) or fraction collector, called the mobilization step. This is achieved by a variety of approaches including salt mobilization or hydraulic mobilization by introducing pressure, vacuum or gravity. It would also be possible to avoid the mobilization step, and perform the detection by capturing an image of the whole capillary with the focused proteins in place. However, any subsequent analysis of the fractions would not be possible in this approach.[66] In order to precisely define the position of pI values in an IEF experiment, usually a set of markers with known pI values are applied. In gel-based IEF, these markers are run in a separate adjacent track along the samples,[67] yet in the solution-based IEF such as CIEF, the markers are spiked in the sample for each individual run.

Hyphenating CIEF with MS, known as analytical CIEF, can be regarded as an advanced variant of the 2-D gel electrophoresis principle; the first dimension being the IEF and the second dimension being the MS analysis (detection of proteins based on mass over charge ratio). However, in the coupling of CIEF to MS, several factors should be considered and resolved. Connecting CIEF to the MS instrument via ESI source is a challenge, due to the cross talk of the electric fields applied in CIEF for separation and the electric field applied in ESI for

ionization.[68] Moreover, it shall be considered that CIEF buffers can contaminate the sample and interfere with MS analysis. The differences in local pH of the eluted peaks from CIEF and the zero net charge of the analytes might also add complication to the ESI performance.

For achieving a deep analysis of complex protein mixtures, CIEF can be used as the first dimension of fractionation, known as preparative CIEF, coupled to orthogonal reverse-phase liquid chromatography, tandem mass spectrometry (RPLC-MS/MS). Preparative CIEF acquires larger amount of sample compared to the analytical CIEF and, therefore, separation columns have larger inner diameters than the analytical ones.[69]

As noted above, the advantage of using CIEF over IPG-IEF methods is the online nature, as well as minimal sample requirement (at nanoliter scale) thanks to the higher sample recovery and the greater availability of miniaturized devices. In addition, advantages include the time- and cost- effectiveness of the technique. However, there is also a problem inherited to CIEF from the IEF concept, which is the analyte (protein or peptide) precipitation and aggregation due to the zero net charge on the surface at the focusing position. This can cause blockage in the capillary and fluctuation in electric current as well as migration time. This issue can partially be addressed by the addition of non-ionic or zwitterionic detergents and surfactants to enhance protein solubility, such as Tween, CHAPS and chaotropic agents like urea, or organic modifiers such as glycerol.[70] Yet, the compatibility of additives with ESI-based MS analysis should be considered, and additional step of sample clean-up to MS analysis should be accounted.

Conclusively, the application of separation and enrichment techniques, and the use of these methods in combination for multi-dimensional separations, allow for deeper analysis of proteomic samples by mass spectrometry that consequently shed light on the holistic view of the studied proteomes.

## Deep Proteomics

A holistic description of the complete human proteome demands the challenging detection of altering forms of proteins in complex biological samples. As discussed in previous sections, analytical methods for covering this complexity often require large sample quantities and multi-dimensional fractionation. These requirements, of course, influence the throughput of the methods. However, one can increase the sensitivity and throughput of protein quantification by limiting the number of analytes such as peptides and proteins in each experiment. Therefore, proteomics experiments are often divided into two categories: discovery and targeted proteomics. Discovery proteomics aims at unbiased “global” identification and quantification of as many proteins as possible to reach the deepest level of the proteome. This step requires spending more time and effort per sample and reducing the number of samples analyzed. Whereas in targeted proteomics strategies, which is in most of the time a follow-up step on the hypotheses generated from the discovery phase, the number of monitored analytes are limited. Also, the pipeline including the separation and instrument tuning are optimized to achieve the highest sensitivity, specificity and throughput for testing the hypotheses on a bigger sample cohort.

Deep proteomics at discovery phase ultimately aims at uncovering all proteoforms unambitiously. For that, in an attempt for MS-based proteomics, direct identification of full-length proteins by MS is called ‘top-down’ approach. The top-down approach offers a rich data for identification and characterization of the protein molecules. Nevertheless, the generated data is often convoluted, thus analysis is challenging. Moreover, limitations in both MS and fractionation instrumentations for top-down proteomics limit the application area. However, these challenges are being more and more addressed in recent years and intact protein analysis is shifting from the study of individual proteins (subject of targeted proteomics),[71,72] towards the deep proteomics approach (aiming at the discovery proteomics).[73]

The dominant approach in proteomics today is called the ‘bottom-up’ approach.[74] This strategy always involves a digestion step that breaks the protein molecules into defined peptides prior to MS analysis. This approach, also known as ‘shotgun proteomics’, avoids many of the problems associated with the top-down approach. Here, the protein identification is inferred by the

detection of the corresponding peptides by the MS/MS.[75] The main challenge in shotgun proteomics is to maximize protein identification and quantification in a cost- and time- efficient manner.[76] The general complication with shotgun proteomics that is still a challenge in proteome analysis is the ambiguity[75] in characterizing all alternative splice forms and modifications as well as endogenous protein cleavages, in addition to the complex combination of these variants.[77]

In MS-based protein identification often only m/z of the protein or peptide alone is not sufficient evidence to identify the precursor ion. Therefore, as previously explained in mass spectrometry section, usually a tandem mass spectrum is required. However, limitation on the speed of the mass spectrometers results in different strategies on tandem MS. The two most common methods for acquiring MS/MS in shotgun proteomics experiments are called data-dependent acquisition (DDA) and data-independent acquisition (DIA). In DDA, from a full-scan mass spectrum (MS1) with mass information on intact peptides, a limited number of precursors are selected for acquisition of fragmentation (MS/MS) spectra.[78] While in DIA, MS/MS scans are collected systematically and independently of the precursor information. DIA is available in various formats such as collecting fragmentation data without precursor-ion selection[79], using ion mobility–collision-induced-dissociation time-of-flight mass spectrometry[80], using wide isolation windows[81] and using narrow isolation windows combined with many injections.[82]

Although both methods have shown their strength and power as versatile strategies, there are also limitations and disadvantages associated with each method. For DDA, the instrument speed of sampling limits the number of peptides that are sampled regardless of the dynamic range and peak capacity of the mass analyzer. In a complex sample, only a few number of precursors, based on their abundance, are selected for MS/MS, which can result in most peptides being unsampled.[83] Moreover, the randomness of the method can result in variation of sampled peptides between replicate measurements.[84] Besides, the precursor abundance of a peptide may stay lower than the background threshold of the MS1 spectrum, regardless of the detection level of MS/MS; or not be picked up at the highest abundance of elution peptide, or fragmented in a complex mixture of chimeric co-eluents.[85] On the other hand, for DIA method, often the search for assignment of peptide sequences for the

obtained MS/MS spectra in databases is less effective than DDA, as a result of the elevated complexity of MS/MS data.

In many cases, the biological difference, which is going to be addressed by proteomics studies, requires not only the identification of proteins in samples but also quantification of the abundances of identified proteins. Mass spectrometry is also associated with difficulties in the ionization efficiency, and detectability of different peptides in a given sample. Therefore, different methods are introduced for both relative and absolute quantification of proteins in samples.[86] The intensity of the ions in a mass spectrum is not proportional to the absolute number of the injected ions into the MS. However, the ratios of the peak intensities of the same analyte between different samples, if all conditions of the injection and analysis are kept constant, can accurately reflect the relative amounts of that analyte in different samples.[8] Based on this concept, one approach for relative quantitation is to separately analyze samples by MS and compare the spectra. In this method, which is usually applied to the results of LC-MS/MS analysis, the corresponding peak in the LC-chromatogram for each ion is extracted for each sample to determine peptide abundance in that sample. Comparison of the area under these peaks in different samples results in a relative abundance of that analyte. This method is known as label-free quantification.[87,88]

Another strategy in relative quantification is based on labeling the analytes in different samples by incorporation of a tag into the analytes, usually composed of different isotopes for different samples. The labeled samples are mixed and analyzed in the same LC-MS/MS experiment. Relative abundance of the corresponding “reporter” peaks in the mass spectrum then represent the relative quantities of the analytes in the samples. This strategy is called label-based quantification. In general, there are two types of labeling methods developed for label-based quantification. One is the stable isotope metabolic labeling and is based on the incorporation of chemically same amino acids with differences in isotopic composition. Resulted proteins as well as digested peptides will have different m/z, which will be visible in mass spectrometry. The relative abundance of different peaks of each isotopic composition represents the relative quantification between those proteins in the corresponding samples. One of the best developed methods for metabolic

isotopic labeling is stable isotope labeling with amino acids in cell culture (SILAC).[89]

The second strategy is called isobaric mass tags, which is labeling of peptides with tags. These tags have similar total masses and chemical structure, but different isotopic compositions in a part of the tag known as mass reporter region. This mass difference, therefore, is balanced in each tag with complement region known as mass normalization region. These two parts together balance the total mass of the tag to be similar for all labels. Each tag is chemically bound to the digested peptides of proteins in each sample. After tagging samples, they can be mixed and analyzed together. Tandem MS on each peptide results in the analysis of the reporter ions and relative abundances of the corresponding peaks represent relative quantities of the protein in the samples. Among these methods are tandem mass tags (TMT)[90,91] and isobaric tags for relative and absolute quantitation (iTRAQ)[92].

Absolute quantification by MS requires calibration by known amounts of a synthetic analyte such as a peptide, in order to calculate the absolute amount of the same analyte based on a relative comparison to the synthetic one. The method, which is known as selected reaction monitoring (SRM) is usually performed by tandem MS, in which the analyte of interest is selected in the first stage and is fragmented followed by detection of the ion products.[93] This method is also applied on a mixture of analytes known as multiple reaction monitoring (MRM), where pre-defined *m/z* ions are chosen for fragmentation and only specific product ions are detected and reported.[94]

Taken together, the emerging techniques in MS-based proteomics, both for the sample preparation and analysis steps, as well as data acquisition and interpretation, is opening the path towards deeper studies of proteomic samples that aim at the proteome-wide analysis of different specimens. Among the practiced proteomic samples, an area with large research interest to perform deep proteomics is the analysis of blood, since it encompasses a perfect reflection from nearly all healthy and diseased tissues of the body. However, the huge dynamic range of protein abundances and the enormous variety of proteins in blood makes such studies a challenge.

## Blood Plasma Proteome

Blood plasma is among the largest collected clinical specimens. It is considered as the most complex human-derived proteome. The complexity arises since blood contains not only plasma proteins, the ones with defined main function in circulation but also proteins that are actively being secreted or even leaked from different tissues into the blood. The list will drastically expand if one considers the proteoforms of these proteins, including splice variants and post-translational modifications. One instance is glycosylation as most blood proteins are known to be heavily glycosylated.

Albumin is the most abundant protein in blood and represents almost 55% of the total protein concentration. The French chemist Pierre Macquer first introduced the term albumin in the year 1777 for substances that coagulate upon heating. It took more than half a century till the 1830s when the German chemists Justus Von Liebig and Gerardus J. Mulder described albumin as a protein. Later, in 1862 the term 'globulin' was introduced by Carl Schmidt for proteins that are insoluble in pure water. Applying electrophoresis, the Swedish Nobel laureate Arne Tiselius separated blood plasma proteins in the year 1937 and classified them as albumin, α-, β-, and γ-globulins. Advances in electrophoresis and other separation methods resulted in further identification and characterization of more and more proteins in the blood.

According to the early definition that was given in the 1970s by F. W. Putnam in his famous book series: The Plasma Proteins: Structure, Function and Genetic Control, the plasma proteins are those that carry out their function in circulation. Noticeably, this definition does not cover the proteins in blood plasma with origins from tissues that play other roles. Examples of such important proteins include cytokines, peptide hormones as messengers between tissues, proteins that are being leaked due to tissue damage, as well as aberrant secretions from tumors. Considering the importance of such proteins, the term blood plasma proteome has recently been re-defined to cover every protein present in plasma that is detectable by protein analytical means. Elaborating the first definition by Putnam, N. L. Anderson and N. G. Anderson introduced a comprehensive classification in their well-recognized article: "The human plasma proteome: history, character and diagnostic prospects". Accordingly, the plasma proteins are classified in one of the following groups:

• **Secreted proteins from solid tissues with a known function in plasma.** This category contains classical plasma proteins that are secreted from liver and intestines. These are mostly larger than molecular cut-off of kidney filtration (45 kDa) and thus show a long residence time in plasma.

• **Immunoglobulins, also referred to as antibodies.** This group of immune proteins is a vastly complex family of functional proteins in plasma with estimated over 10 million different amino acid sequences in adult human circulation.

• **Long distance receptor ligands.** This group of proteins includes classical peptide and protein hormones such as insulin and erythropoietin. These proteins appear in a wide range of molecular sizes, which regulates their life span in plasma.

• **Local receptor ligands.** This group of proteins comprises of cytokines and other local mediators with small molecular size and, therefore, a short residence time in plasma.

• **Temporary passengers.** This group encompasses the non-hormone proteins such as lysosomal proteins that are actively passing through plasma from their secretion source to their functional site.

• **Tissue leakage products.** These proteins are mostly intracellular proteins, yet can be released into plasma due to cell damage or cell death. Important diagnostic markers, such as myoglobin as a biomarker of myocardial infarction, are often described in this group.

• **Aberrant secretions.** This group contains proteins that are released from tumors or other diseased tissues and includes cancer markers.

• **Foreign proteins.** This group of proteins includes alien molecules with origin from infectious organisms or parasites that are released into the plasma.

The quantitative dynamic range of plasma proteome is limited by the existing state of analytical methods in proteomics. Serum albumin with the normal concentration of 35-50 mg/mL is at the high abundant end, whereas interleukins and cytokines such as interleukin 6 (IL-6) with the normal range of 0-5 pg/mL is at the low abundant end. Thus, the normal plasma proteome has a dynamic range of at least 10 orders of magnitude. This means that the detection

of an analyte such as IL-6 in plasma among the albumin molecules is equivalent to looking for one particular molecule in a crowd of over 7 billions of other molecules.

Nevertheless, despite the vast dynamic range, proteins in all levels are used as biomarkers. Plasma proteins at the very high abundance level such as serum albumin have been described as a marker of severe liver disease or malnutrition. Tissue leakage proteins at the middle abundance level such as cardiac myoglobin are known as a marker for myocardial infarction. And cytokines at the low abundance end such as interleukin 6 are characterized as markers of inflammation or infection.

The path in plasma protein and proteome analysis that was started by Arne Tiselius applying electrophoresis, has been continued by 2-D gel electrophoresis by Anderson, and was later emerged by MS analysis. Recent advances in mass spectrometry techniques opened a new era in blood plasma proteomics. However, the dynamic range of the most advanced mass spectrometers is still in range the of 4 to 5 orders of magnitude. Therefore, the large dynamic range of protein concentrations in blood is currently the greatest challenge for the MS-based plasma proteomics.[95] Thus, many scientific efforts have been focused on the development of methods in order to address this issue.

Pre-fractionation of samples prior to mass spectrometry analysis using conventional separation techniques, such as chromatography and electrophoresis, can increase the depth of proteome analysis by reducing the complexity of samples. These methods are used as the first dimension of separation, either for proteins or after digestion into peptides, prior to LC-MS/MS analysis.[95,96] Furthermore, immunoaffinity-based depletion of highly abundant proteins has become the dominant approach in blood plasma proteomics to reduce the dynamic range of blood samples. Several kits and methods are available for plasma depletion via a single or multi-step procedure.[97-102] However, besides an increase in time and cost of sample preparation, depletion can cause an unwanted removal of non-targeted proteins or result in elevated variability of protein abundances.[103,104] Alternatively, target proteins of interest can be enriched in blood plasma samples. For instance, lectin columns can be applied for the enrichment of glycosylated polypeptides.[35,105] In addition, several cell types release large (microparticles - 100 nm to 1 µm) or small size (exosomes - 40 to100 nm) vesicles into the blood

that contain cell-specific proteins.[106,107] Isolation of microparticles and exosomes is used to enrich for tissue leakage proteins.[108]

Biomarker discovery in plasma proteome is of especial interest due to the existing routine clinical sampling, but also for the fact that blood plasma contains the protein markers that represent the state of the body at a given time. Particular interest is in plasma proteins in the middle abundance range generated from tissue leakage, from which a pathological state can be identified. This range of abundances is accessible for MS-based proteomics studies. Therefore, intensive research is ongoing both for biomarker discovery and pathological study of diseases, in particular when the target organ is not easily accessible, such as neurodegenerative diseases of the brain.

## Alzheimer Disease

In 1901, Alois Alzheimer interviewed a patient and reported a series of symptoms that distinguished the studied disease from what was clinically known. The patient was a 51-year-old woman who was diagnosed with memory decline and cognitive disabilities. Alois Alzheimer and his colleagues defined the clinical and pathological details of this late-life mental decline syndrome.[109,110] The disorder was named Alzheimer disease (AD) after Alois Alzheimer and later in the 1960s it was described as the most common form of senile dementia.[111,112] Currently, it is estimated that more than 27 million people live with AD and the number is predicted to increase to 86 million people by the year 2050. Age is the highest risk factor for AD and its incidence increases exponentially with aging, in which it doubles every 5 years after the age of 65.[113]

The gradual and progressive decline in two or more cognitive domains, most commonly involving executive functions and episodic memory are the hallmark clinical phenotypes of AD.[114,115] These declines are therefore sufficient to cause severe social or occupational impairments. AD dementia is progressive and associated with functional impairment, whereas the cognitive changes of aging are benign and relatively static. This is the main clinical distinction between cognitive changes of aging and those of underlying dementia.[115-117]

In terms of neuropathological characterizations, the hallmarks of AD are more or less similar to what Alois described in his original work, the presence of “miliary bodies” that we now know as amyloid plaques and “dense bundles of fibrils” that we today call neurofibrillary tangles. In 1985, the amyloid plaque was first purified and analysis of its core identified the 4 kDa amyloid β (Aβ) peptide as the main component of the plaques.[118] Consequently, further studies on the protein led to the cloning of the gene encoding the amyloid precursor protein (APP).[119] On the other hand, parallel studies on the neurofibrillary tangles revealed abnormally hyperphosphorylated forms of the protein tau.[120] These attempts were the start of modern research on AD that provides detailed knowledge to understand the APP metabolism and Aβ generation, as well as tau homeostasis. Two important proteins that participate in APP metabolism are prsenilin proteins. Mutations in APP or one of the presenilin genes (PSEN1 or PSEN2) have been reported to cause rare familial forms of AD.[121] However, these mutations are the cause for less than 1% of AD cases, and only 10% of AD patients younger than 65 are with a family history of AD. Therefore, AD is believed to be an age-dependent sporadic disease rather than a genetic disorder.[122] A well-known theory called the ‘amyloid cascade hypothesis’ emphasizes the particular role of Aβ and its imbalance (production vs. clearance) as the driving force of AD. Accumulation of the Aβ then leads to tau pathology and causes neurological degradation and ultimately dementia.[123] However, for late-onset AD, this hypothesis has not yet been proven with adequate certainty.

There are several other hypotheses for the cause of AD, including the “tau hypothesis” that considers hyperphosphorylation of tau proteins as the causative effect for AD,[124,125] the “cholinergic hypothesis” that states the deficiency in production of acetylcholine that is a vital neurotransmitter as the initiative cause of AD,[126] the “oxidative stress hypothesis” which emphasizes the role of reactive oxygen species (ROS) and free radicals as a cause of AD,[127] and the “protein aging hypothesis” that relates protein degradation, such spontaneous chemical modification of asparaginyl residues, as a cause of AD.[128,129]

Considering any hypothesis that explains the AD onset, the long-term preclinical pathology behind AD suggests that disease-modifying drugs will most probably be more effective in earliest stages of AD.[130-132] Therefore, it is of crucial importance to diagnose AD in earlier stages, before the start of

dementia (prodromal AD) or even before the occurrence of symptoms (preclinical AD).[133] In the preclinical stage of AD, the disease progressive mechanism has already been started in the brain; however, the severity of neuronal damage is not yet sufficient to cause any detectable cognitive symptom. In the prodromal stage of AD, known as the mild cognitive impairment (MCI), the causal pathology of the disease will affect the cognition. The diagnosis of AD with certainty has not yet been addressed in these early stages. To find a therapeutic agent that would affect the main disease hallmarks and positively revert the disease progress is still a challenge.[134] Thus, identifying reliable biomarkers to enable early AD diagnostics is currently an obvious requirement of the field.

# Chapter 2
# Present Investigations

# Main Objectives

The aim of the current thesis is to develop a novel technique for online fractionation of peptides and proteins according to their isoelectric point (pI), in order to perform deep MS-based shotgun proteomics studies. The main focus is on the instrumentation development and optimization, as well as clinical applications in biomarker discovery. The 7.4 hypothesis is presented, which assumes that in blood samples of patients with neurodegenerative disorders, there is a higher ratio of abnormal to normal proteins, in a range of 7.30 to 7.50 of the pI distribution of proteins. This hypothesis is tested for biomarker discovery in early-stage AD.

# Specific Aims

In more detail, the specific aims of the current thesis, as presented in five scientific publications (**Papers I-V**), are the following:

1. Setting up the pI fractionator device for separation of polypeptides based on their isoelectric point in electrospray-friendly solution.

2. Optimizing the pI fractionator device and separation conditions for application in deep shotgun proteomics.

3. Optimizing and setting up a pipeline for high content shotgun proteomics by LC-MS/MS.

4. Setting up an online desalinator and couple it to the pI frationator device for applications in blood plasma proteomics.

5. Applying the instrument for investigation of protein/peptide levels in Alzheimer disease (AD) plasma samples, particularly at the region around pH = 7.4.

# Methodology

## Blood Samples

In the current thesis, blood samples were analyzed by deep shotgun proteomics for potential predictive markers of AD. The samples were collected from two cohorts: Kuopio blood plasma cohort, acquired within the EU project PredictAD [135] in Finland, and DemVest plasma cohort, acquired in Norway.

The Kuopio biobank consists of samples including patients at different AD stages of stable and progressive mild cognitive impairment as well as healthy controls. Respective data is available on gender and age. In the current study, twelve progressive mild cognitive impairment (PMCI) samples and the equal number of healthy controls with matching age and gender were pooled into two respective samples to be further examined.

Serum samples were also recruited from an ongoing cohort study of patients with mild dementia in Western Norway (DemVest cohort). Selection and diagnostic procedures were performed as previously described.[136] Data is available on the rate of annual decline, which is calculated based on the five years follow-up measurements of the mini-mental state examination (MMSE) score. Fourteen female samples from donors at the age of 77 ± 6 years were selected, of which six samples have had the annual decline rate in MMSE score of below 2 (slow decliners), and 8 samples have had the annual decline rate of over 5 (fast decliners).

## Ethical Permits

Regarding Kuopio blood plasma cohort (studied in **Paper IV**), the use of samples for biomarker discovery has been approved by the research ethics committee, Hospital District of Northern Savo, stating the research fulfils the Finnish Research law (488/1999), Finnish personal data act (523/1999) and Finish Rights for Patients law (785/1992).

Concerning DemVest cohort (studied in **Paper V**), blood serum was sampled and stored at University of Bergen, Norway. The use of samples has been approved by the ethical committee at University of Bergen (reference number: 167.04) and additionally approved by the ethical committee at

Karolinska Institutet, the regional human ethics committee of the Stockholm County (ethical permit number: 2013/914-31/4).

## The MJ-CIEF Device

A fractionator device, called multiple-junction capillary isoelectric focusing (MJ-CIEF), is developed for in-solution separation of proteins and peptides, applicable for online coupled mass-spectrometry analysis.[137] PEEK capillary is utilized (OD 635 µm, ID 395 µm) as seven equal segments of 1.55 cm long, joined via insertions of 0.2 cm tubular Nafion membranes (OD 610 µm, ID 330 µm). The assembled capillary is passed through eight 0.5 mL Eppendorf tubes for supplementation of external electrolytic solutions. Compared to the traditional fused silica capillary columns, the PEEK material shows no electroosmotic force; therefore, the analytes are separated merely based on their charge (pI), rather than size-to-charge ratio. Moreover, the PEEK material is also inert in terms of adsorbing polypeptides, due to hydrophilic property of the surface. The multi-compartment assembly of the developed tool provides the ability to stabilize pH gradient via permeable membranes at each junction, which results in low sample contamination by carrier ampholytes and well-focused zones.

In order to ensure a linear electrical field gradient across the entire column, a voltage divider was implemented in the device. The multi-section voltage divider was assembled by connecting eight resistors (1 MΩ, 1 W) in series, with the last one connected to the voltage supplier polar. In addition, the joint section of each two neighboring resistors was connected to interval vials via an immersed platinum wire (Figure 3).

A stepwise focusing and mobilization scheme was implemented for eluting each of the focused zones. After the first focusing step, hydrodynamic flow was applied for elution of the peaks from the anode side. Then, the cathode electrode was moved forward to the next junction and refocusing was done in the reduced volume. The process was repeated, and the anode fraction was collected sequentially. The developed optimized method for refocusing enables fractionation with increased resolution and has the ability of rapid release of the zones by higher flow rate.

MJ-CIEF was used as a micro-preparative fractionator device for separation of complex peptide mixtures based on pI in **paper II** as well as blood plasma protein separation in **papers IV and V**.

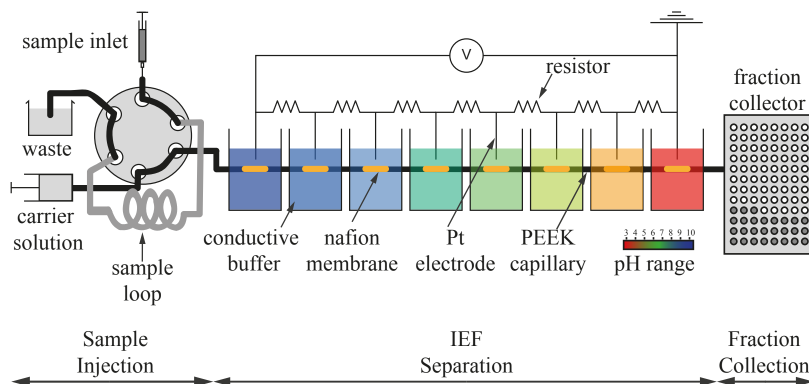


**Figure 3**. **Scheme of the MJ-CIEF device**. The device is comprised of a 6-port valve, connected to the multijunction IEF column consisting of PEEK capillaries with Nafion membrane windows immersed in vials. The vials contain solutions with different pH values. Resistors of the voltage divider ensure stable and linear distribution of the electrical field along the column.

## Fractionation Workflow

External electrolyte buffer solutions are used to provide the pH range of pH=3 to pH=10 through the anodic vial of the MJ-CIEF device to the cathodic vial (i.e., acetic acid, ammonium acetate, ammonium formate, ammonium bicarbonate and ammonia). 5 % isopropanol and 0.5 % Pharmalyte 3-10 in miliQ water is applied as carrier buffer. First, the sample (3 µL) is loaded into a loop, connected to a 6-port injector valve. Through adjustment of the valve, the sample can be transferred into the fractionator. High voltage (-1 kV, -100 V/cm) in a constant current regime, limited to maximum 25 µA, is applied to the end vials. The ground electrode is immersed in the collector end (anode), and the negative voltage is applied at the injector end (cathode). The focusing is normally done in 25 min and -0.2 kV voltage is applied during mobilization. The focused fractions are collected at 0.5 µL per min flow rate. After collection of each focused fraction, the mobilization is stopped for performing the refocusing. The cathode electrode is moved forward in the adjacent vial. The vial buffer is altered to ammonia, and the high voltage is applied for 5 min (-100 kV/cm) for refocusing. The mobilization-refocusing cycle is repeated for all vials sequentially.

## Desalinator

In sample preparation for MS analysis, sample contamination by salts and buffers as well as sample loss in preparation steps seem inevitable.[138] Protein extraction and digestion utilize salts such as dodecyl sulfate, deoxycholate, urea, dithiothreitol and iodoacetamide. Proteins (or peptides) separation methods also incorporate salts such as ionic solutions in strong cationic/anionic exchange, or ampholyte mixtures during isoelectric focusing, which are not compatible with mass spectrometry.[137-140] Fractionation of blood plasma proteins by MJ-CIEF device is hindered by the high salt concentration and electrical conductivity of the plasma, which interferes with the electrophoresis process. Hence, addition of a sample clean-up procedure by a desalinator can play an important role in the improvement of the sample preparation workflow.

Standard available methods for desalination and buffer exchange are centrifugal filtration (size exclusion) and solid phase extraction (SPE). Filtration needs costly filters, as well as addition of time-consuming centrifugal steps to the procedure. The sample recovery and filter cut-off specificity, as well as buffer and filter compatibility, are also considerable issues. In SPE, target molecules attach to the surface of the stationary phase, while salts and buffers are washed away.[139] Afterward, target molecules are eluted from the stationary phase in a solvent that is generally compatible with downstream steps, such as 0.1% trifluoroacetic acid (TFA) or formic acid (FA) in water. Many approaches use a C4 coated submicrometer-porous silica beads for the purification of proteins and C18 beads for peptides. Despite many advantages of these methods, they are still time-consuming, labour-intensive and relatively expensive. Moreover, robustness and sensitivity are issues for small amounts of protein, due to inefficient interactions and low recoveries.[140] Suitable molecular cut-off (less than 400 Da, as measured and reported in **Paper IV**), chemical and thermal stability of Nafion membrane, and availability at submicron diameters, make it suitable for microdialysis performance. Therefore, we applied Nafion membrane to develop a micro-chamber for sample clean-up and buffer exchange, as a complementary desalinator section of the MJ-CIEF device. The microdialysis device has a similar configuration to a cell that has previously been reported for hydrogen/deuterium exchange.[141] The desalinator consists of two coaxial tubes (4 cm). The inner tube (390 µm ID and 356 µm OD) contains the sample and is separated from the outer tube by a Nafion membrane.

Washing buffer is flowing in the outer tube (2 mm ID), and the sample is desalted before it enters into the fractionator section (Figure 4).

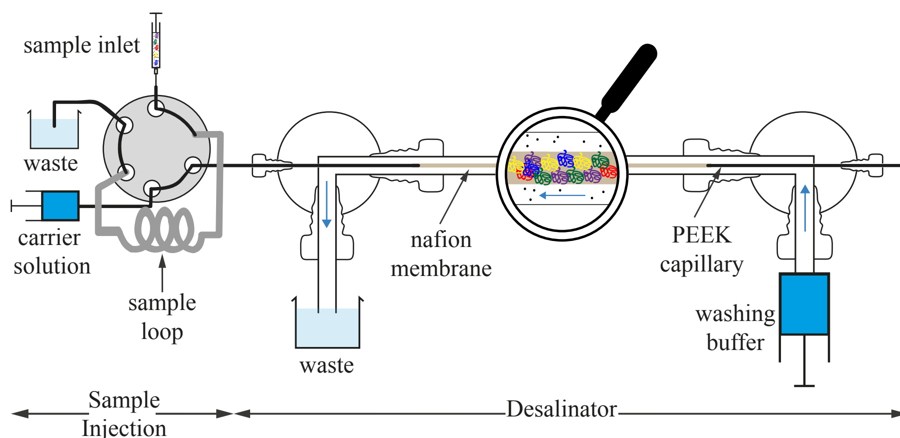


**Figure 4**. **Scheme of the desalinator device**. Sample is injected via 6-port valve and passes through a PEEK capillary tube into the Nafion membrane. The washing buffer passes through the outer tube that results in buffer exchange of the sample.

## Protein Digestion

Proteins, either from cell lysate (as described in **Paper III**) or from blood plasma/serum (as studied in **Papers IV and V**) were digested in solution.[142] Proteins were reduced and alkylated with DTT and IAA at the final concentration of 10 mM. Proteins were digested with sequencing grade trypsin at 37 °C overnight. All peptide mixtures were purified using C-18 columns.

## LC-MS/MS

Samples were analyzed in a random order, using LC-MS/MS method. In brief, an EASY-Spray LC column (PepMap® RSLC, C18 material with 100 Å pores, 50 cm in **paper III**, and 15-cm column in **papers II, IV and V**) from Thermo Scientific was applied with a flow rate of 250 nL/min. The gradient was nonlinear and was optimized to have constant rate of peptide elution from LC column. Mass spectra were acquired in a data-dependent manner using a

top-10 MS/MS method with an Orbitrap XL (**paper I**), Velos (**paper II**), Q Exactive (**paper III**) or Q Exactive Plus (**papers IV and V**) mass spectrometers from Thermo Fisher Scientific. MS1 spectra were acquired at a resolution of 70,000 with a target value of 3E+06 ions or a maximum accumulation time of 250 ms in the m/z range from 300 to 2000 for plasma proteins as reported in **papers IV and V**, and 400 to 1200 for cancer cell lysate as described in **papers II and III**. MS/MS spectra were acquired using CID (**paper II**) or otherwise HCD fragmentation at a resolution of 17,500 with a target value of 2E+05 ions or a maximum accumulation time of 120 ms.

## Data analysis

Acquired data from LC-MS/MS is in the form of a .raw file containing all MS1 and MS2 spectra. There are several available search engines for identification of peptides and parental proteins from these files. Moreover, the identified proteins should be quantified precisely for a study based on the abundance of proteins. There are also well-developed software packages that are available for quantification. In the current thesis work, we have used MaxQuant for peptide and protein identification and quantification.[87] As a sequence database, the International Protein Index was used. Modern Orbitrap mass spectrometers can provide up to ppb (parts per billion) level of mass accuracy[87] in both MS1 and analytic MS/MS scans. For Orbitrap Q-Exactive instruments, mass tolerances are commonly set as 10 ppm for precursor (MS) and 20 ppm for fragments (MS/MS). Cysteine carbamidomethylation was selected as a fixed modification, and oxidation of methionine, N-acetylation of protein, and deamidation of asparagine and glutamine were selected as variable modifications. The results were filtered to a 1% false discovery rate (FDR) at both protein and peptide levels.[143] MaxQuant reports several .txt files as the results of identification and quantification analysis. Further data analysis was performed on the proteinGroups.txt, which contains the relative abundance of protein groups. If the proteome of a sample is fractionated, the software is reporting the total abundance of each protein prior the fractionation. The MaxQuant-reported 'LFQ-intensity' of each protein was taken as relative protein abundance. It is also possible to introduce each fraction as a separate experiment, and consequently the software will report the abundance of each protein in each fraction. By means of this information, the focusing position of

each protein was determined by calculating the weighted average of all values (Figure 5). Statistical tests and calculations were done using R and SIMCA14.0 (Umetrics, Sweden). Orthogonal projections to latent structures discriminate analysis (OPLS-DA) was performed using SIMCA.[144] For that, LFQ-intensities were log transformed. The performance of the resulted model was reported as the cumulative correlation (R2X[cum]), and predictive power was reported based on seven-fold cross-validation (Q2[cum]).

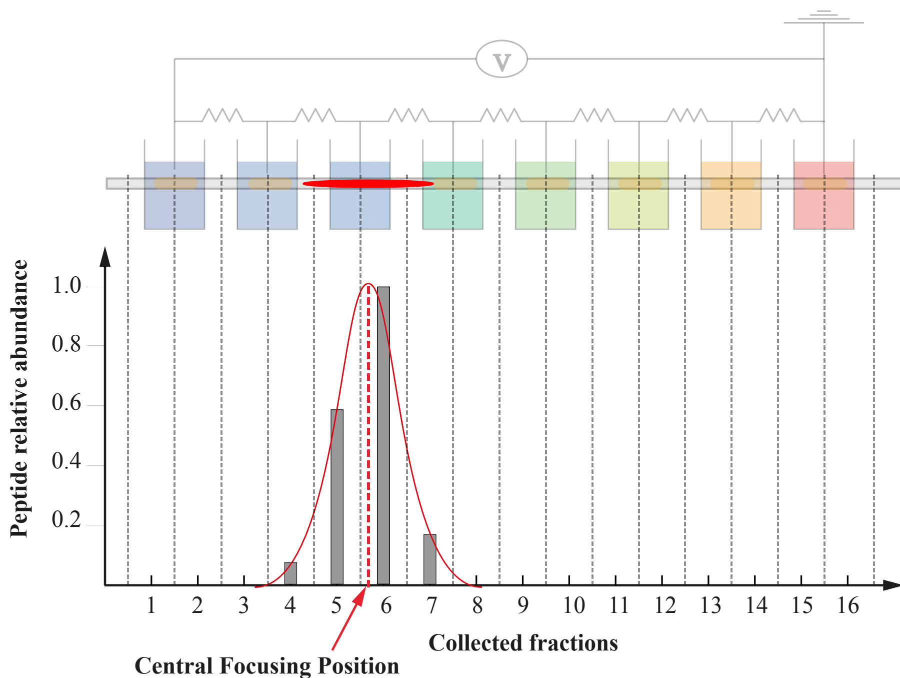


**Figure 5**. **Calculating the focusing position of each peptide**. Each peptide has its focusing position in the CIEF column. For a peptide that is shown in red, during fraction collection, it appears in several of collected fractions close to its central focusing position. LC-MS/MS analysis of fractions reviles relative abundance of the peptide in each fraction. The central focusing position is calculated by fitting a bell-shaped curve to the distribution of peptide abundances and determination of the central position of the curve, or calculated the weighted average of all values.

# Results and discussion

LC-MS/MS has become a dominant analytical tool in current human proteomics, providing most suitable integral performance for the comprehensive analysis of high-complexity protein and peptide mixtures extracted from tissues, cell lines and body fluids. Nevertheless, the complexity and large dynamic range of protein concentrations are among the greatest challenges for the MS-based proteomics studies.

Sample fractionation prior to LC-MS/MS analysis is currently generally accepted as a key to cover proteome of complex samples to a reasonable depth. Separation of proteins and peptides by isoelectric point is one of the emerging techniques for this purpose. The advantage of using isoelectric focusing for pre-fractionation, which is an orthogonal separation to RPLC, is not only in the reduction of sample complexity, but also in grouping polypeptides by their isoelectric point (pI). The pI-value is an informative parameter that is associated with proteins amino acid sequence, structure, and post-translational modifications.

This thesis work includes the development of a novel device for online peptide and protein fractionation according to their isoelectric point for deep shotgun proteomics. For that purpose, a novel multiple-junction capillary isoelectric focusing fractionator (MJ-CIEF) was developed and subsequently improved regarding the resolving power and reproducibility of fractionation. In addition, a novel algorithm was developed to calculate the identified peptides' pI and distinguish PTMs based on the elution time. Moreover, an online desalinator was developed and was coupled to the device for direct sample buffer-exchange and isoelectric separation of human blood plasma.

To achieve the aim of deep proteomics, a multi-parameter optimization of an LC-MS/MS pipeline was performed. The optimized experimental conditions included the optimization of cell lysis and protein extraction, digestion of insoluble cell debris, tailoring the LC gradient profile and choosing the optimal dynamic exclusion window in data-dependent MS/MS as well as the optimal m/z scan window.

Conclusively, the combination of developed desalinator, pI fractionator device and optimized deep proteomics LC-MS/MS method was applied in biomarker discovery for AD in blood plasma, in which a panel of new potential

biomarkers is introduced. By means of the pI information of proteins, the 7.4 hypothesis, i.e. a higher ratio of abnormal to normal proteins in the range from 7.30 to 7.50 of the pI distribution of proteins, was also tested, and its potential as a biomarker for early diagnosis of AD is investigated.

All results of the current work are summarized in five scientific papers as explained in the following pages.

## Paper I- Developing an online multiple-junction capillary isoelectric focusing fractionator (OMJ-CIEF) device

An online multiple-junction capillary isoelectric focusing fractionator (OMJ-CIEF) device was developed for the separation of proteins and peptides in solution by pI. The separation column was constructed of seven equal sections, joined via tubular Nafion membrane insertions. Each junction was merged into a reservoir of the external electrolytic buffer, which was used both to supply electrical contact and for solvent exchange. The performance of the fractionator was explored using protein and peptide samples covering a broad pI range from pH= 3 to 10.

Separation was carried in ionic and ampholytic buffers, including ammonium formate, ammonium hydroxide, histidine and arginine. For that, all reservoir chambers and separation column were filled with the same buffer. The pH of the external buffer was adjusted to increase linearly from 2.9 in the anodic end, up to 8.5 in the cathodic end.

Five measurements were performed by injecting myoglobin into OMJ-CIEF coupled to mass spectrometry. In these experiments, myoglobin was eluted after 0, 10, 35, 80 and 180 minutes by applying an electrical field. Comparing the eluted peak by time indicates the migration of myoglobin towards the anodic side as well as focusing.

Separation of a mixture of three proteins, myoglobin, cytochrome c and lysozyme in ammonium hydroxide buffer was carried over 80 minutes. Peak shape of released separated proteins indicated peak broadening of later fractions. To confirm that the peak broadening occurs during mobilization, the same experiment was performed, but with switched polarity of the electric field. Lysozyme, which was the last and broadest peak in the previous experiment, was now first and sharpest peak. This result confirmed the peak broadening during mobilization. To circumvent this problem, selective release of downstream analyte fractions was achieved by maintaining electric potential across upstream segments of the capillary after the focusing stage.

Using single-component ampholyte buffers such as histidine with pI = 7.3 and arginine with pI = 10.8, controlled separation of protein mixture into basic and acidic fractions was demonstrated. Proteins with higher pIs than the applied ampholyte buffer were focused at the cathodic end within the high

pH range, and proteins with pIs lower than ampholyte buffer were observed at the anodic end of the column within the low pH region.

Separation of tryptic peptides from BSA digest was also performed in 5 mM aqueous histidine both as a carrier solution and electrolytic liquid in external reservoirs. IEF was carried over 1 h in 5 μA constant current. Focused peaks were mobilized for ESI-MS detection via hydrodynamic pumping at 3 μL/min. The peak elution time of identified peptides (37 peptide covering 60 % of BSA protein sequence) were compared to the theoretical pI values. The negatively charged peptides with (pI < 5) were drifted to the anodic end with low pH (equivalent to tR ≈ 1.5 min), whereas positively charged peptides (pI > 7.6) were drifted towards the cathodic end with the higher pH region (late released peaks with tR ≈ 5 min). Moreover, an intermediate fraction of peptides with pI from 6 to 7 was also focused at the middle region of the column (tR ≈ 3 min).

## Paper II- Optimizing the MJ-CIEF device as a micro-preparative fractionator

Although the presence of a reasonably high concentration of ampholytes is not pleasant for the LC-MS analysis due to increased background and ion suppression effect, it is important for the CIEF process, to establish a stable pH gradient and thus improve the separation performance.

Different concentrations of pharmalytes (one of the commercial carrier ampholytes widely used for isoelectric focusing experiments) from 0% as the control to 0.25, 0.5, 1, 2.5 and 5% were spiked in a mixture of the yeast proteome. LC-MS/MS experiments were performed on each sample in three replicates. The data was analyzed, and the number of identified peptides were compared along the increase in pharmalyte concentration. The identified peptides showed a significant decrease by adding pharmalyte in concentrations higher
than 0.5 %.

Compensating the stability of the pH gradient during IEF in lower concentrations of ampholytes, we used the advantage of multi-junction structural compartment of the MJ-CIEF device, by applying a range of buffers, from pH=3 in the first vial, increasing by one unit of pH as going towards the last vial that ends up to pH=10. This strategy made it possible to maintain a stable pH gradient in low concentrations of pharmalyte (0.5%) during focusing. The collected fractions were further diluted six times before injection into RPLC, and thus the final Pharmalyte concentration in the LC−MS/MS run was below 0.1%.

Electrical field provides the driving force in isoelectric focusing experiment by moving the charged compounds towards the field until they reach the region where pH = pI and they are neutralized. Therefore, it is essential to keep the electrical field linear and strong during the CIEF experiment. Beside the analyte and running buffer with proton and hydroxyl ions that maintain the conductivity,[145] there are other ions and salts that accumulate on anode and cathode poles along the CIEF experiment time. Also, different analytes have different mobility and focusing speed in a mixture. Early focused analytes can create nonconductive local regions in the CIEF column (known as hot spots) and, therefore, interfere with focusing of other components.

To maintain the linearity of electric field, a voltage divider was designed and added to the MJ-CIEF structure. The effect of the performance of MJ-CIEF device was studied with two experiments on separation of digested yeast proteome. In one experiment, IEF separation was performed with MJ-CIEF and the second one with MJ-CIEF equipped with voltage divider. The focusing position of each peptide was calculated by fitting the abundances of that peptide in different fractions into a bell-shaped curve. The center of the curve was determined and reported as the focusing position of that peptide. Focusing position of all peptides in each experiment was calibrated to the theoretical pI values using a polynomial of fifth order, and the deviation from the calibration curve for each peptide was measured. Standard deviation from distribution of these residual values σ represented the quality of separation and performance of the device. The σ value for MJ-CIEF was 0.435 without the voltage divider and 0.230 with the voltage divider, which was significantly improved. This result demonstrates that the addition of voltage divider improves the performance of MJ-CIEF fractionation.

In addition, as previously noted, the mobilization of focused zones without peak broadening is a challenge in CIEF. Peaks are getting broader by traveling along the column during elution. To overcome this problem, we applied a stepwise refocusing and releasing of the peaks from the column. After the focusing step, the peaks are eluted from the column as much as one junction on the column. Remaining peaks are subjected to refocusing again within the remaining part of the column. This refocusing sharpens back the peaks, particularly the very last basic peaks with very high pI values, which is a challenge in IEF. To study the effect of refocusing on performance of the MJ-CIEF, the same sample of digested yeast proteome was again separated, but this time peaks were collected with stepwise mobilization and refocusing. The σ value decreased into 0.21, and more peptides were identified compared to direct elution of peaks (Figure 6).

The acquired .raw files from the last experiment were also searched for modified peptides, deamidation of Asn and phosphorylation on Ser, Thr or Tyr. The same calculation based on the abundances of the modified peptides in different fractions was performed to calculate the central focusing position of each peptide. This value was later compared with the focusing position of unmodified form of the peptide, and the shift was calculated. The obtained

results indicate a different pI shift for different peptides, in a way that the shift for the same PTM is not constant and depends on the sequence of the peptide. Therefore, the pI values and pI shifts (ΔpI) due to PTM represent a sequence-specific information that can be used in reducing the false discovery rate in shotgun proteomics.

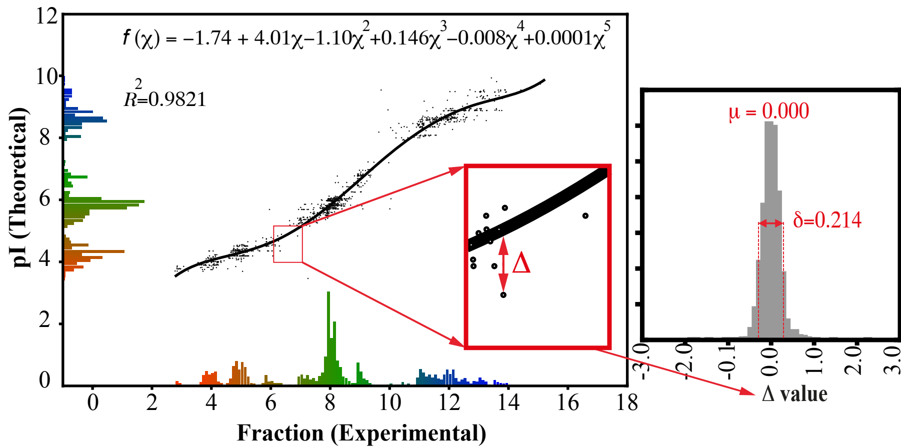


**Figure 6**. **Focusing position of peptides versus theoretical pI.** Focusing position of peptides are determined, and theoretical values are calculated. The results are plotted, and a polynomial curve was fitted. σ value is the standard deviation of the pIs from the fitted curve.

## Paper III- Rapid and deep single-dimension shotgun proteomics

Urea, SDC and ProteasMax are three commonly used detergents for cell lysis and protein extraction. These detergents were used in a comparative experiment for best efficacy, where extracted samples were analyzed by tryptic digestion followed by LC-MS analysis. The total number of identified proteins and peptides were compared, and the use of ProteasMax resulted in highest values. GO annotation analysis of identified proteins shows higher number of proteins in membrane, nucleus and cytosolic fractions. ProteasMax detergent (Sodium 3-((1-(furan-2-yl)undecyloxy)carbonylamino) propane-1-sulfonate) is a hydrophobic anionic sulfonate. It is sensitive to temperature and pH, in which it degrades in acidic conditions. Therefore, it is rather easy to remove it after protein digestion and prior to LC-MS analysis of peptides. Addition of 10% v/v acetonitrile to the ProteaMax lysis buffer also facilitates solubilization of hydrophobic proteins.

There are several organelles and cell debris, which are insoluble in cell lysis buffers. These compartments of the cell are usually removed from cell extract prior to protein digestion. However, the pellet contains proteins, which are low soluble, or trapped in cell debris. To investigate the increase of the proteomics depth, the precipitated organelles were maintained in lysate buffer during digestion. This was in a comparison with removed precipitated organelles from lysis buffer during the digestion by centrifugation. Results of LC-MS/MS analysis of both samples indicated 7% higher protein identification by keeping the cell debris. GO annotation also indicated that the extra gained proteins are most notably from the nucleus and membranes.

In conventional linear gradient on RPLC, the rate of peptide elution is not constant. This is causing saturation of MS analysis capacity at some regions and low efficacy in others during an LC-MS/MS analysis. To optimize the efficacy of an LC-MS/MS run and to make a constant elution rate of peptides from LC column, the gradient was experimentally tailored to a nonlinear profile. The increase in the identified number of peptides would result from the peptides that were eluted in later stages of the run instead of the saturated zone. The optimized profile for a 4 hours experiment was set from 2% acetonitrile to 5% over 19 min, and then to 19% over 133 min, and finally to 30% over 28 min and finally followed by a 20 minutes washing step. This profile resulted in

identification of 35,623 unique peptides, which showed an increase of 2000 peptides over the linear gradient.

In data-depended analysis mode, each full MS scan is followed by a specified number of MS/MS of the most abundant peptide ions detected by full MS. To minimize multiple analysis of the same precursor ion, analyzed precursors are automatically uploaded to a dynamic exclusion list for a certain time. High sensitivity of mass spectrometers such as Q Exactive results in peak picking at very early steps of elution of the LC column, in which the signal is not strong enough for a full informative MS/MS analysis. To overcome this issue, by means of very fast scanning Q Exactive instrument, in a series of 3 hours gradient runs, the DEW was decreased in intervals including 5, 15, 30, and 90 seconds. Results suggested the DEW of 15 seconds as the optimum value, in which the over selecting of the same peak (high abundance precursors) still did not decrease the ion selection of low abundant ions. It should also be considered that this value is dependent to the gradient and speed of the peptides elution from the LC column.

Considering the range of m/z distribution of tryptic peptides in a shotgun proteomics analysis, the scan range for best informative peptides was limited from 400 to 1200, where the default scan range for MS1 spectra in Q Exactive is from 300 to 1650. In a comparison experiment of these two values, the limited range of 400 to 1200 resulted in identification of 159 more protein groups (up to 4044 protein groups). Normally the short sequence peptides have poor sequence specificity and are not informative for protein identification. On the other hand, the very long peptides often result in uninformative MS/MS spectra because of the insufficient fragmentation. Therefore, limiting the scan range, results in focusing on sequencing of more informative peptides as well as excluding most of the background ions such as solvent and salt clusters, LC stationary-phase components and ubiquitous contaminants (e.g. siloxanes or phthalic acids).[146]

The time of gradient and length of the LC column are directly proportional to the resolving power of RPLC separation. Based on the above-mentioned optimized parameters, including cell lysis and digestion, nonlinear LC gradient profile and mass spectrometry parameters, a series of experiments including four time-points, 2, 3, 4 and 5 hours gradients with 3 LC columns, 15 cm, 25 cm and 50 cm length were performed. Using 50 cm LC column in 4-h LC-

MS/MS run with a 3 h gradient time 37,554 peptides and 4,825 protein groups were identified. Merging three replicates increased the identified peptides to 56,390 and 5,354 protein groups, which represent nearly half of the human proteome.[76] Further increase in the gradient time results in slow increase and gradually reaching a plateau in identification number. Therefore, the increase in gradient time becomes rather inefficient in terms of the content rate (i.e. the number of proteins identified per unit time) (Figure 7).

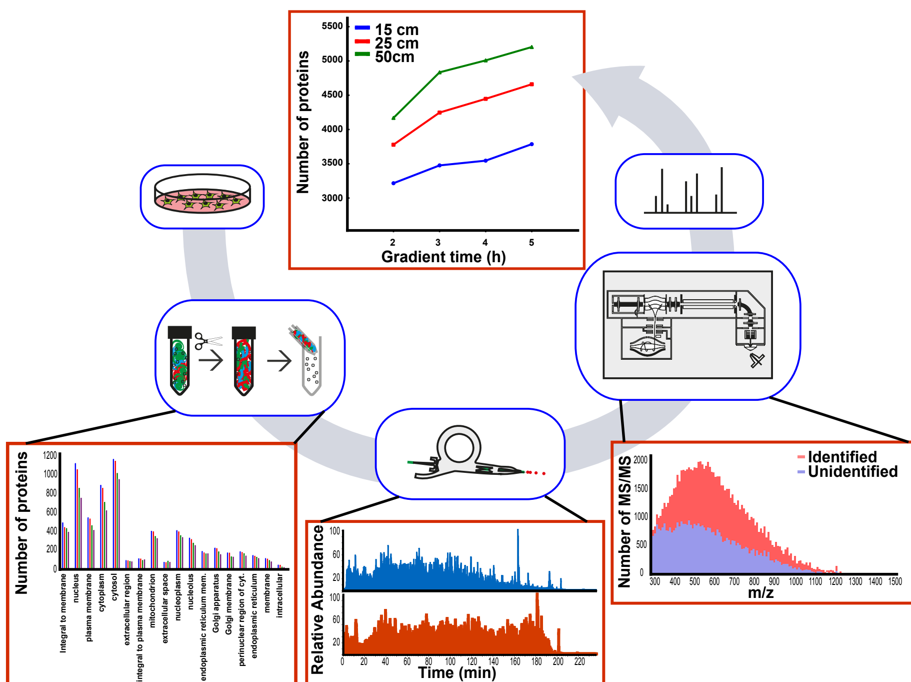


**Figure 7**. ProteaseMAX buffer was found to be more efficient for cell lysis and protein extraction compared to SDC and urea. Preserving cell debris in cell buffer for protein digestion enabled higher coverage of membrane and nucleus proteins. Tailoring the LC gradient profile yielded higher total number of sequenced proteolytic peptides. The optimal dynamic exclusion window setting in data-dependent MS/MS was found to be 15 s for 3 h LC gradient time, and the optimal scan window was 400-1200. Comparison between three different column lengths and also the gradient time illustrated the gain in identification number by longer column, but saturation in very long gradient runs.

## Paper IV- Online desalination and isoelectric focusing of plasma proteome

Blood plasma contains high concentration of salts and metabolites that are conductive and, therefore, interfere with electrophoresis. An online microdialysis desalinator was developed using Nafion membrane. The molecular cut-off of this membrane was determined to be around 400 Da. This is less than the size of proteins and neuropeptides but large enough for removing salts and small metabolites. The constant flow on the outer layer of the membrane increases the buffer exchange rate.

In the performed experimentation, 365 protein groups from 4030 peptides were quantified with 1% FDR at both peptide and protein levels. In a separate experiment, unfractionated blood plasma was analyzed in ten replicates that resulted in quantification of 206 protein groups from 2571 peptides. Comparison between these two sets of the results indicates the increase in depth of the proteome due to pre-fractionation. Moreover, the correlations between replicates of each group were calculated, where the median value of Pearson correlation was 0.96, and the CV was 3.8%, representing the high reproducibility of the method.

Furthermore, the normalized protein abundances were compared between two sample groups, i.e. healthy controls and AD patients. Some of the proteins showed statistically significant different abundances. Alpha-2-macroglobulin (A2M) is a proteinase inhibitor, and one of the high abundant proteins in blood plasma and the presence of this protein was previously shown in amyloid plaques.[147] This protein is also a potential marker for the damage in blood-brain barrier, which can be present in neurodegenerative diseases such as AD.[148] Moreover, it is suggested that A2M plays a role in regulation of the immune response, the association of which is reported with AD.[149,150] Plasmin is another high abundant protein in blood and is shown to promote Aβ peptide clearance.[151] Four members of complement system proteins showed elevated levels in AD patients. There are studies that suggest the activation of this system by soluble Aβ peptides in the AD brain.[152,153]

Additionally, for each sample, by means of the abundances of each protein in all fractions, the centroid of the focusing position for that protein was calculated. The obtained focusing positions of all proteins were compared

between replicates, which resulted in median CV of 3.5%. This low value of CV shows the accuracy of measuring this parameter experimentally. Furthermore, comparison of focusing position of each protein between healthy and patient samples showed statistically significant shift for some proteins such as kinesin-like protein, thyroxine-binding globulin and Msx2-interacting protein (Figure 8). The change in focusing position indicates the change in pI, and this could be due to many reasons such as altered splice variance, protein truncation, protein-protein interactions or PTM. In most of the cases, the observed shift was towards acidic pH. Some important PTMs cause acidic pI shift such as deamidation and phosphorylation. Kinesin-like protein and thyroxine-binding globulin has several positions for phosphorylation, but this was not observed in identified peptides of this study. However, for Msx2-interacting protein several sites of Asn deamidations were identified in samples from AD patients. Asn deamidation leads to formation of the damaging isoaspartyl (isoAsp) residues and the elevated levels of isoAsp residues in Msx2-interacting protein was shown earlier in blood plasma with AD progression.[129]

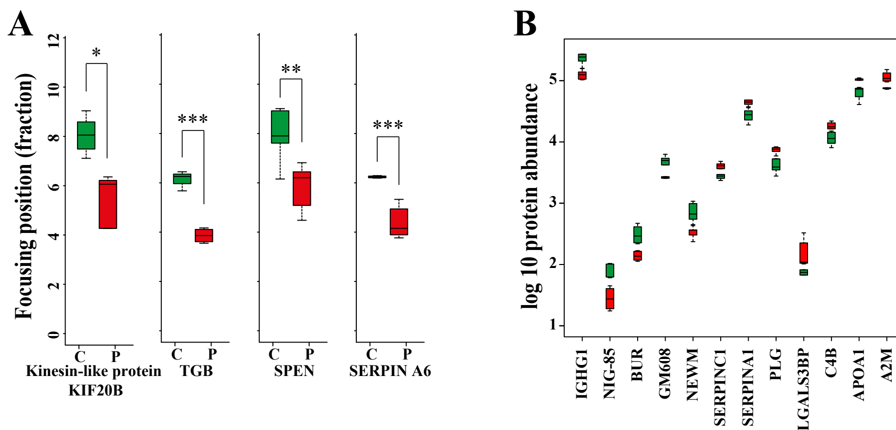


**Figure 8**. Proteins with **(A)** significant differences in their pI values or **(B)** protein abundances, between healthy controls (green) and AD patients (red).

## Paper V- Narrow pH range around 7.4 in biomarker discovery of neurodegenerative disorders

For each of the 14 serum samples (6 patients with slow memory decline and 8 patients with fast memory decline), separation was performed. For each sample, 20 fractions were collected with MJ-CIEF, digested and analyzed by LC-MS/MS. The subsequent data analysis resulted in quantification of 650 protein groups with false discovery rate less than 1% at both peptide and protein levels.

Pearson correlation between quantified abundances of proteins between samples in each group was higher than r = 0.96 between the replicates, which shows high reproducibility of the experiments. Comparison of proteins abundances between two patient groups showed statistically significant changes (with p-value less than 0.05) for several proteins. Some of the proteins have been reported in previous studies, in which the abundance change is associated with the progression of neurodegenerative diseases. Clusterin, a chaperon protein, was detected in amyloid plaques that can play a role in degradation of Aβ peptide.[154,155] Beta-Ala-His dipeptidase has a role in regulation of immune response and was reported as a shared biomarker candidate for multiple sclerosis (MS), Alzheimer disease (AD), and Parkinson disease (PD). The level of this protein is higher in AD and PD patients compared to healthy controls, but lower in MS patients.

pI fractionation of proteins results in focusing of each protein in a fraction with the closest pH to its pI. Analyzing the abundance of each protein in different fractions revealed the focusing position of that protein in each sample. By means of spiked pI-markers, all proteins with 7.1 < pI < 7.5 in each sample were selected. Sum of the proteins abundances in each fraction was calculated. As a control comparison, the same calculation was performed for proteins in neighboring fractions in which the pI markers were well identified in all samples. The protein concentrations in the fractions were compared between the replicates of two groups of patients, i.e. slow memory decliner vs. fast memory decliner. This revealed significantly (p-value = 0.003) higher concentration of total protein in fast decliner patients near the 7.4 region. Therefore, the obtained result of this comparison indicates strong potential of the protein concentration at the pI=7.4 region to be used as a potential biomarker for progressive AD.

Identified proteins in this region are also of important interest. Further study regarding the abundances of these proteins in an unsupervised clustering resulted in perfect separation of the two groups of patients. Some of these proteins were also reported in previous studies as markers of AD. Apolipoprotein C-I, calmodulin-like protein 5 and caspase-14 are among those proteins that were present in the fraction with pI≈7.4 in blood serum of patients with fast memory decline but not in patients with slow memory decline.

Using OPLS-DA approach, two predictive models were built. The first one was based on the results of proteins abundances from label-free quantification, and the second one was based on proteins abundances with pI in the 7.4 region. An OPLS-DA model based on proteins abundances was built with R2=0.9 and Q2=62%, whereas the model based on proteins in the pI≈7.4 region had R2=0.9 and Q2=72%. This result indicates excellent predictive power of both models, with a higher predictive power of the second model compared to the first one (Figure 9).

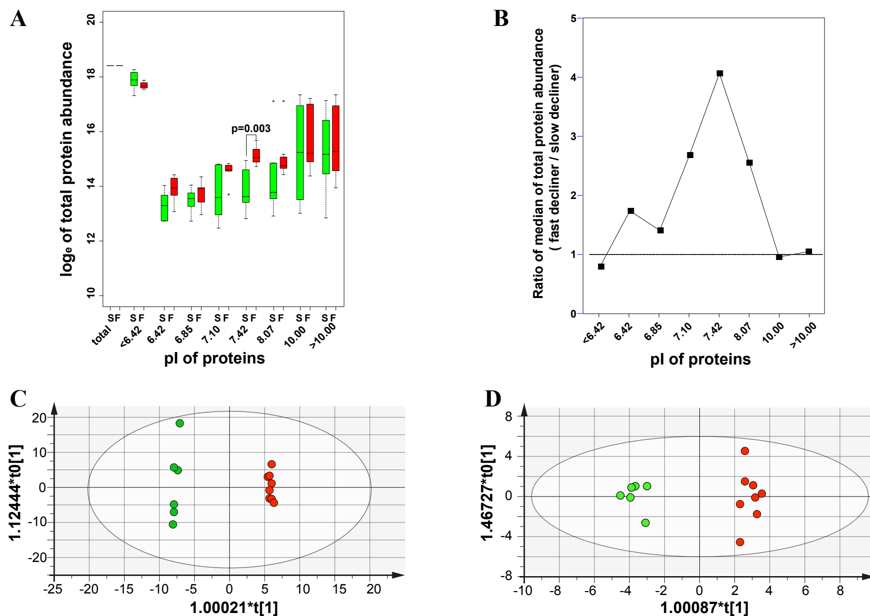


**Figure 9**. **(A)** Protein concentration in different fractions. Green boxplots represent slow decliners, and red ones show the fast decliners. **(B)** Each dot is the ratio of median protein abundances in all fast decliners to the slow decliners. **(C)** OPLS-DA model is shown based on total protein abundances and **(D)** based on proteins abundances in the 7.4 region. Green circles represent slow decliners, and red ones correspond to the fast decliners.

# Chapter 3
# Conclusion and Future Perspectives

# Conclusion and Future Perspectives

Fast growing applications of deep proteomics in cell biology and medical studies demand the development of advanced techniques with higher sensitivity, throughput, lower cost and compatibility with automation. Mass spectrometry is the core technique of proteomics studies. This necessitates inventing novel techniques, as well as improving available upstream methods compatible with this instrument. Isoelectric focusing is one of the most widely applied methods in separation of polypeptides prior to mass spectrometry analysis. In the current thesis, an in-solution isoelectric focusing separation of proteins and peptides is presented for applications in MS-based clinical proteomics.

We developed a novel micropreparative capillary isoelectric focusing device. The utilized porous membrane allows the online coupling of the device to auto-sampler, fraction collector, detector or mass analyzer. The pH gradient and chemical profile inside the column can be tailored externally via Nafion membranes in each junction. This significantly decreases the need for ampholyte mixtures for creating and maintaining the pH gradient in separation column, and minimizes the contamination of the sample and instrument by salts, buffers, and ampholytes. Separation column is made of PEEK capillary that precludes the electroosmotic flow to interfere with separation. The higher loading capacity compared to analytical capillary IEF enabled the application of the device as first dimension of separation prior RPLC hyphenated to MS. Orthogonality of IEF separation to RPLC is also another merit.

Addition of a voltage divider to the MJ-CIEF stabilizes the linear electric field along the separation column, and increases the effective length of the column. As an advantage, the issue of interfering hot spot regions, i.e. local nonconductive regions, as well as high electrical conductivity of anode and cathode ends of the column were addressed. Mobilization of focused peaks in CIEF is often associated with peak broadening. The multi-junction structure of the MJ-CIEF allows the stepwise elution of peaks from one junction and refocusing of the remaining ones.

Furthermore, an online microdialysis device was developed for direct desalination and buffer exchange of samples, particularly blood plasma, prior CIEF. The same Nafion membrane was applied as a dialysis membrane with

molecular cut-off of ca. 400 Da. This molecular cut-off allows removing of salts, amino acids and metabolites from blood plasma proteins and peptides, which are conductive and interfere with electrical field during IEF separation.

In order to achieve the highest efficiency in deep shotgun proteomics, the LC-MS/MS pipeline was also optimized without any hardware or software modification of the commercial instrument. This optimized pipeline complied with the developed MJ-CIEF fractionation method for a two-dimensional deep shotgun proteomics approach.

The application of MJ-CIEF in prefractionation of peptides prior to LC-MS/MS analysis was also demonstrated. The separation pipeline is fast, not labor-intensive and cost-effective. The results prove the reproducibility and robustness of the pipeline in deep proteomics, representing via high correlation between replicates and low coefficient of variation (CV). MJ-CIEF coupled to desalinator was also applied in pre-fractionation of blood plasma proteins followed by shotgun proteomics. Two-dimensional separation by IEF-LC increased the depth of the plasma proteome analysis. This results in identification of several proteins as potential biomarkers for AD, in which the abundance of protein was significantly changed. Applying the MJ-CIEF in fractionation of proteins provided additional information of pI. The isoelectric point of a protein depends on the sequence, structure, PTM and protein-protein interaction, hence is biologically informative. Several proteins with significant shift in their pI value were identified in biomarker discovery for AD patients. These altered proteins are also introduced as additional potential biomarkers for AD.

The pIs of normal blood plasma/serum proteins avoid the values around 7.4 that is equal to the pH of blood. This makes the region attractive for proteomics-based biomarker discovery. The proteomics study of the narrow pH range around 7.4 suggests significant elevation in total protein abundance in AD patients with fast memory decline. Moreover, the results suggest the accumulation of several proteins in the 7.4 region in AD patients with progressive memory decline. These proteins and the total protein abundance at 7.4 region can be potential markers for AD progression. The introduced biomarkers are at the discovery phase, and the investigation was a proof-of-concept test made on a small cohort. Further confirmation of these results on a bigger cohort of diverse gender and age, as well as the specificity of these

biomarkers to AD compared to other neurodegenerative diseases, will be considered to validate the clinical application of the presented biomarker candidates.

Development and optimization of the MJ-CIEF device and the coupled desalinator in this thesis work is going to be continued by commercializing the device to be available in more advanced and automated shape for laboratories who are practicing proteomics. Automation of the device will make it more user-friendly as well as high-throughput. Microfluidic chip design of the separation column will be considered to increase the robustness and reproducibility of the device performance. Regarding the data analysis step, the obtained pI values of the separated peptides and proteins present additional important information. This can be implemented in the search engines for identification of protein and peptides and reducing the false discovery rate.

The developed instrument is also highly applicable in studying PTMs, particularly the ones that induce a shift in pI of the molecule. For that, more investigation on these PTMs and the association between the pI shift and the sequence of the polypeptide is needed. Shotgun proteomics lacks vital information on the proteoforms, and application of the pI shift in biomarker discovery requires the identification of the particular proteoform. This is possible with merging the pI separation with top-down proteomics.

Taken together, the outcomes of scientific experiments along this thesis signify the valued applications of both the developed MJ-CIEF instrument and investigation of the narrow pH range around 7.4 of the body fluids in proteomics research. This will ultimately lead us to a deep and holistic understanding of biological processes and disease mechanisms, as well as discovery of biomarkers that enable the prevention, diagnosis and effective treatment of neurodegenerative disorders.

# Acknowledgment

# Acknowledgment

I would like to acknowledge my supervisors, teachers, collaborators and colleagues, friends and family, from whom I got the best of support and company all along the fantastic Ph.D. journey.

My sincere thanks go to my principal supervisor **Roman A. Zubarev**. Thanks for the opportunity you gave me to do a Ph.D. in your lab and countless support along this journey. I am truly grateful for your instructions that advanced my abilities in scientific methods, critical thinking and applying the power of statistics. You provided a nice, friendly and multicultural environment for us where I gained experience in teamwork with people from different scientific backgrounds. Your enthusiasm to the science, especially less explored areas, always inspired me. I also thank you for your kind feedback on this thesis work.

I would like to express my gratitude to my supportive co-supervisors. **Dag Årsland**, thank you for giving me support and guidance along the projects, particularly when it comes to neurology and brain. Many thanks for sharing your interesting science and insights with me. **Thomas Helleday**, please accept my gratitude for your kind support along this journey. I always appreciated your punctuality and prompt feedbacks.

I am also especially grateful to **Thorleif Lavold**, the CEO of BioMotif AB, our industrial company partner. I had a great time working with your company and enjoyed our communications and trips we had around the world. You made the work a joy. Thanks for introducing me to the industrial and business partners of the scientific community.

I would like to especially acknowledge my mentor, **Juan Astorga-Wells**, a big thank you for all the practical tips and the best of support you provided me along the way. Thanks for being there for me whenever I needed your support. Thanks for being such an affectionate and supportive colleague. I also appreciate your feedback on this thesis.

I wish to acknowledge **Steven Gygi** from the Department of Cell Biology, Harvard Medical School, for giving me the valued opportunity to do an internship in his lab. Thanks for sharing your intriguing science with me.

Everyone in Gygi Lab, especialy **Ryan Kunz**, I am grateful for the best friendly and supportive environment you provided for me throughout my visit.

I would like to thank all my co-authors and collaborators for their fruitful teamwork in projects. Many thanks for being part of this work.

**Peter Roepstorff**, I truly enjoyed our scientific meetings, starting at the early phases of the projects. It has been inspiring and fruitful working with your group.

I would also like to offer my thankfulness to **Konstantin Chingin**. Thanks for the moments we enjoyed discussing all the details of the OMJ-CIEF device and thanks for all the inspiration you offered me at the start of the journey. Also thanks for the proofread of this thesis.

I am also very much grateful to **Bo Zhang**. Thanks buddy for your valued company along the journey. I wish you all the best, in your future life and research. I always enjoyed chatting about our projects and the Ph.D. journey.

It is my pleasure to acknowledge all the PIs and senior researchers at the Department of Medical Biochemistry and Biophysics (MBB) for sharing their great scientific achievements with us. I would like to particularly thank **Hans Jörnvall**, **Craig Wheelock**, **Tomas Bergman**, **Josefin Skogsberg** and **Jimmy Ytterberg** for all their supports within these years.

And I wish to thank my former and current colleagues at Zubarev Lab. **David Good**, thank you for the encouraging early days when I joined the lab. **Alexey Chernobrovkin**, it was always fascinating to discuss with you on the projects and new ideas. **Amirata Saei Dibavar** and **Marjan Abbasi**, thanks for great times during the collaborative works and all nice chats and lunches. **Harshavardhan Budamgunta**, thanks for productive teamwork we had together. **Hongqian Yang**, thanks for being such a kind colleague. **Xueshu Xie** and **Nataliya Tarasova**, we joined the lab almost at the same time, and it was wonderful to experience the Ph.D. journey with you.

**Carina Palmberg**, **Marie Ståhlberg** and **Dorothea Rutishauser**, thanks for your continuous assist and all the back-up with the instruments and reagents. **Susanna Lundström**, I enjoyed the interesting talks around IgG-omics. Also thanks for final proofread of the Swedish abstract. **Shiva Kalantari**, thanks for sharing your scientific enthusiasm with us during

your visit. and I am also grateful to your kind feedback on the Persian abstract. **Marta Guerrero Valero**, **Consuelo Marin Vicente**, **Ernesto Gonzalez**, **Alexandra Bernadotte**, **Aleksandr Manoilov**, **Hao Piliang**, **Neil Rumachik**, **Muna Muse**, **Tina Heyder**, **Pierre Sabatier**, **Oleksii Rebrov**, **Atim Enyenihi**, **Liban Abikar** and **Yaroslav Lyutvinskiy**, please accept my appreciation for being part of the group, for all the interesting and memorable scientific and social events.

I would also like to thank the members of Wheelock Lab, especially **Antonio Checa**, **Anas Kamleh** and **Cristina Gomez**. I want to give my gratitude to **Ella Cederlund** from the Jörnvall Lab for all the instrumentation support and **Walid Tajeddinn** from Årsland Lab for kindly being collaborative and supportive. **Taraneh Shirazi** and **Mahsa Ebrahimi** from Schüler Lab, thanks for all great times at the department and during the lunchtime.

I am also truly grateful to our kind administrative support **Victoria Balabanova** and our great IT support **Chad Tunell**, as well as the administrator of postgraduate education **Alessandra Nanni**. Thanks for your everyday assist. I would also like to thank **Svante Backlund** for all the time and technical support when I needed to use the workshop, tools and machines at the MBB department.

No work will find its way to success without the generous financial support. I wish to acknowledge the Alzheimersfonden, the Swedish Research Council (Vetenskapsrådet), VINNOVA Foundation and the Knut and Alice Wallenberg Foundation. In particular, I appreciate the acceptance of VINNOVA Foundation (grant within the EuroStar) to support my work in continuation of the project towards commercializing the MJ-CIEF device. I would also like to thank the Karolinska Institutet travel grant, Hirsch travel grant and IMSF travel grant for their financial support during scientific visits and conferences that gave me the unique opportunities to visit many labs and fantastic researchers around the world.

Last but not least, I would like to thank all my friends, who made this journey unforgettable. **Hassan Foroughi Asl** and **Afsaneh Kasiri**, **Iman Kamali Sarvestani** and **Nazgol Shahbandi**, **Mohsen Karimi**, **Atosa Estekizadeh** and **Parnian Karimi**, **Siamak A. Kamranvar**,

**Laia Sadeghi** and **Soheil A. Kamranvar**, **Alireza Azimi**, **Mona Azimi** and **Soudabeh Rad**, thank you all for wonderful time and joyful moments we all had together.

I am forever indebted to my mother, **Fatemeh Karbasi**, thank you for all the countless support. My sisters **Atefeh Pirmoradian** and **Mina Pirmoradian**, thanks for your limitless love and I wish you the best of achievements in your studies and life. My fabulous grandmother **Hamideh Khanom Nasrlolahi** and my supportive uncles and aunts, particularly **Mohammad Karbasi**, and **Rababeh Karbasi** and your wonderful families, please accept my sincere thanks. I wish to also thank my mother-in-law **Fereshteh Manhoubi**, thanks for all you support and love. A very warm and special thanks to you, my kind brother and sisters in-law and their fabulous families.

And finally my most special thanks go to my beloved wife **Mahya Dezfouli**. Thanks for your endless support, care and kindness. Thanks for creating so many beautiful moments in our life. And thanks for your unremitting encouragement in this journey.

# Bibliography

# Bibliography


1 Poste, G. Bring on the biomarkers. *Nature* **469**, 156-157, doi:10.1038/469156a (2011).

2 Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5**, 821-834, doi:10.1038/nrd2132 (2006).

3 Hartley, H. Origin of the word 'protein'. *Nature* **168**, 244 (1951).

4 Ouzounis, C. A., Coulson, R. M., Enright, A. J., Kunin, V. & Pereira-Leal, J. B. Classification schemes for protein structure and function. *Nat Rev Genet* **4**, 508-519, doi:10.1038/nrg1113 (2003).

5 Snider, N. T. & Omary, M. B. Post-translational modifications of intermediate filament proteins: mechanisms and functions. *Nat Rev Mol Cell Biol* **15**, 163-177, doi:10.1038/nrm3753 (2014).

6 Smith, L. M., Kelleher, N. L. & Consortium for Top Down, P. Proteoform: a single term describing protein complexity. *Nat Methods* **10**, 186-187, doi:10.1038/nmeth.2369 (2013).

7 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).

8 Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207, doi:10.1038/nature01511 (2003).

9 O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021 (1975).

10 Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol Cell* **49**, 583-590, doi:10.1016/j.molcel.2013.01.029 (2013).

11 Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575-581, doi:10.1038/nature13302 (2014).

12 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587, doi:10.1038/nature13319 (2014).

13 Thomson, J. J. Cathode Rays. *Philosophical magazine* **44**, 293-316 (1897).

14 Stephens, W. E. pulsed mass spectrometer with time dispersion. *Physical reviews*, 69-691 (1946).

15 H. Sommer, H. A. T., and J. A. Hipple. Measurement of e/m by cyclotron resonance. *Physical reviews* **82**, 697-702 (1951).

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

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

18 Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. *Anal Chem* **85**, 5288-5296, doi:10.1021/ac4001223 (2013).

19 Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711, doi:10.1038/nrm1468 (2004).

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

21 Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**, 709-712, doi:10.1038/nmeth1060 (2007).

22 Zubarev, R. K., NL.; McLafferty, FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.,* **120**, 3265-3266 (1998).

23 Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533, doi:10.1073/pnas.0402700101 (2004).

24 Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-867 (2002).

25 Forner, F. *et al.* Proteome differences between brown and white fat mitochondria reveal specialized metabolic functions. *Cell Metab* **10**, 324-335, doi:10.1016/j.cmet.2009.08.014 (2009).

26 Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol* **189**, 739-754, doi:10.1083/jcb.200911091 (2010).

27 Vermeulen, M., Hubner, N. C. & Mann, M. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **19**, 331-337, doi:10.1016/j.copbio.2008.06.001 (2008).

28 Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631, doi:10.1073/pnas.0812099106 (2009).

29 Mittler, G., Butter, F. & Mann, M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res* **19**, 284-293, doi:10.1101/gr.081711.108 (2009).

30 Grimsrud, P. A., Swaney, D. L., Wenger, C. D., Beauchene, N. A. & Coon, J. J. Phosphoproteomics for the masses. *ACS Chem Biol* **5**, 105-119, doi:10.1021/cb900277e (2010).

31 Macek, B., Mann, M. & Olsen, J. V. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* **49**, 199-221, doi:10.1146/annurev.pharmtox.011008.145606 (2009).

32 Thingholm, T. E., Jensen, O. N. & Larsen, M. R. Analytical strategies for phosphoproteomics. *Proteomics* **9**, 1451-1468, doi:10.1002/pmic.200800454 (2009).

33 Zhong, J., Molina, H. & Pandey, A. Phosphoproteomics. *Curr Protoc Protein Sci* **Chapter 24**, Unit 24 24, doi:10.1002/0471140864.ps2404s50 (2007).

34 Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840, doi:10.1126/science.1175371 (2009).

35 Zielinska, D. F., Gnad, F., Wisniewski, J. R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897-907, doi:10.1016/j.cell.2010.04.012 (2010).

36 Matic, I. *et al.* In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* **7**, 132-144, doi:10.1074/mcp.M700173-MCP200 (2008).

37 Krokhin, O. V. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal Chem* **78**, 7785-7795, doi:10.1021/ac060777w (2006).

38 Moruz, L., Pichler, P., Stranzl, T., Mechtler, K. & Kall, L. Optimized nonlinear gradients for reversed-phase liquid chromatography in shotgun proteomics. *Anal Chem* **85**, 7777-7785, doi:10.1021/ac401145q (2013).

39 Kocher, T., Swart, R. & Mechtler, K. Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Anal Chem* **83**, 2699-2704, doi:10.1021/ac103243t (2011).

40 Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858 (1996).

41 Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469, doi:10.1038/379466a0 (1996).

42 Washburn, M. P., Wolters, D. & Yates, J. R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-247, doi:10.1038/85686 (2001).

43 Branca, R. M. *et al.* HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods* **11**, 59-62, doi:10.1038/nmeth.2732 (2014).

44 LJ., H. Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *Am J Physiol.* **21**, 173-179 (1908).

45 KA., H. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. *Biochemische Zeitschrift.* **78**, 112-144 (1917).

46 Hubner, N. C., Ren, S. & Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* **8**, 4862-4872, doi:10.1002/pmic.200800351 (2008).

47 Wu, S. *et al.* Multi-modality of pI distribution in whole proteome. *Proteomics* **6**, 449-455, doi:10.1002/pmic.200500221 (2006).

48 Norman D. Scott, T. S. MEASUREMENTS OF THE MOBILITY OF EGG ALBUMIN AT DIFFERENT ACIDITIES. *J. Am. Chem. Soc.,* **46**, 2700–2707 (1924).

49 Tiselius, A. *The Moving Boundary Method of Studying the Electrophoresis of Proteins, inaugural dissertation* 7 thesis, (1930).

50 Tiselius, A. A new apparatus for electrophoretic analysis of colloidal mixtures. *Trans. Faraday Soc.* **33**, 524-531 (1937).

51 Tiselius, A. Reflections from both sides of the counter. *Annu Rev Biochem* **37**, 1-24, doi:10.1146/annurev.bi.37.070168.000245 (1968).

52 Righetti, P. G. *Handbook of Isoelectric Focusing and Proteomics*. Vol. 7 xvii–xxii (2005).

53 Righetti, P. G., Sebastiano, R. & Citterio, A. Capillary electrophoresis and isoelectric focusing in peptide and protein analysis. *Proteomics* **13**, 325-340, doi:10.1002/pmic.201200378 (2013).

54 Lee, J. E. *et al.* A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J Am Soc Mass Spectrom* **20**, 2183-2191, doi:10.1016/j.jasms.2009.08.001 (2009).

55 Hjerten, S. Free zone electrophoresis. *Chromatogr Rev* **9**, 122-219 (1967).

56 Jorgenson, J. W. & Lukacs, K. D. Capillary zone electrophoresis. *Science* **222**, 266-272 (1983).

57 Heller, C. *Analysis of Nucleic Acids by Capillary Electrophoresis. Vieweg, Braunschweig*. Vol. 1 pp. 1–313 (1997).

58 Schultz, C. L. & Moini, M. Analysis of underivatized amino acids and their D/L-enantiomers by sheathless capillary electrophoresis/electrospray ionization mass spectrometry. *Anal Chem* **75**, 1508-1513 (2003).

59 Bello, M. S., Capelli, L. & Righetti, P. G. Dependence of the electroosmotic mobility on the applied electric field and its reproducibility in capillary electrophoresis. *J Chromatogr A* **684**, 311-322, doi:10.1016/0021-9673(94)00545-1 (1994).

60 Svensson, H. *Electrophoresis by the moving boundary method. A theoretical and experimental study.*, (Alrnqvist & Wiksells Bok., 1946).

61 Svensson, H. *Arch. Biochem. Biophys., (Suppl. 1*, 132-140 (1962).

62 Svensson, H. *Acta Chem. Scand.* **15**, 325-341 (1961).

63 Svensson, H. *Acta Chem. Scand.* **16**, 456-466 (1961).

64 Vesterberg. *Acta Chem. Scand.* **23**, 2653-2666 (1969).

65 Bjellqvist, B. *et al.* Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**, 317-339 (1982).

66 Wehr, T. *Handbook of Isoelectric Focusing and Proteomics*. Vol. 7 181–210 (Elsevier, 2005).

67 Shimura, K., Kamiya, K., Matsumoto, H. & Kasai, K. Fluorescence-labeled peptide pI markers for capillary isoelectric focusing. *Anal Chem* **74**, 1046-1053 (2002).

68 Zhang, C. X. *et al.* Stepwise mobilization of focused proteins in capillary isoelectric focusing mass spectrometry. *Anal Chem* **72**, 1462-1468 (2000).

69 Camilleri, P., Okafo, G. N., Southan, C. & Brown, R. Analytical and micropreparative capillary electrophoresis of the peptides from calcitonin. *Anal Biochem* **198**, 36-42 (1991).

70 Conti, M., Galassi, M., Bossi, A., Righetti, P. G. Capillary isoelectric focusing: the problem of protein solubility. *J. Chromatogr. A* **757**, 237–245 (1997).

71 Boyne, M. T., 2nd, Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. Precise characterization of human histones in the H2A gene family by top down mass spectrometry. *J Proteome Res* **5**, 248-253, doi:10.1021/pr050269n (2006).

72 Ge, Y., Rybakova, I. N., Xu, Q. & Moss, R. L. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc Natl Acad Sci U S A* **106**, 12658-12663, doi:10.1073/pnas.0813369106 (2009).

73 Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254-258, doi:10.1038/nature10575 (2011).

74 Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362, doi:10.1038/nmeth.1322 (2009).

75 Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4**, 1419-1440, doi:10.1074/mcp.R500012-MCP200 (2005).

76 Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* **13**, 723-726, doi:10.1002/pmic.201200451 (2013).

77 Schluter, H., Apweiler, R., Holzhutter, H. G. & Jungblut, P. R. Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J* **3**, 11, doi:10.1186/1752-153X-3-11 (2009).

78 Stahl, D. C., Swiderek, K. M., Davis, M. T. & Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *J Am Soc Mass Spectrom* **7**, 532-540, doi:10.1016/1044-0305(96)00057-8 (1996).

79 Purvine, S., Eppel, J. T., Yi, E. C. & Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3**, 847-850, doi:10.1002/pmic.200300362 (2003).

80 Myung, S. *et al.* Development of high-sensitivity ion trap ion mobility spectrometry time-of-flight techniques: a high-throughput nano-LC-IMS-TOF separation of peptides arising from a Drosophila protein extract. *Anal Chem* **75**, 5137-5145 (2003).

81 Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods* **1**, 39-45, doi:10.1038/nmeth705 (2004).

82 Panchaud, A. *et al.* Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal Chem* **81**, 6481-6488, doi:10.1021/ac900888s (2009).

83 Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* **10**, 1785-1793, doi:10.1021/pr101060v (2011).

84 Liu, H., Sadygov, R. G. & Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201, doi:10.1021/ac0498563 (2004).

85 Zhang, B., Pirmoradian, M., Chernobrovkin, A. & Zubarev, R. A. DeMix workflow for efficient identification of cofragmented peptides in high resolution data-dependent tandem mass spectrometry. *Mol Cell Proteomics* **13**, 3211-3223, doi:10.1074/mcp.O114.038877 (2014).

86 Ong, S. E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **1**, 252-262, doi:10.1038/nchembio736 (2005).

87 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

88 Lyutvinskiy, Y., Yang, H., Rutishauser, D. & Zubarev, R. A. In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. *Mol Cell Proteomics* **12**, 2324-2331, doi:10.1074/mcp.O112.023804 (2013).

89 Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

90 Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).

91 McAlister, G. C. *et al.* Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* **84**, 7469-7478, doi:10.1021/ac301572t (2012).

92 Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169, doi:10.1074/mcp.M400129-MCP200 (2004).

93 Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9**, 555-566, doi:10.1038/nmeth.2015 (2012).

94 Mitchell, P. Proteomics retrenches. *Nat Biotechnol* **28**, 665-670, doi:10.1038/nbt0710-665 (2010).

95 Pernemalm, M. & Lehtio, J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert review of proteomics* **11**, 431-448, doi:10.1586/14789450.2014.901157 (2014).

96 Gautam, P., Nair, S. C., Ramamoorthy, K., Swamy, C. V. & Nagaraj, R. Analysis of human blood plasma proteome from ten healthy volunteers from Indian population. *PloS one* **8**, e72584, doi:10.1371/journal.pone.0072584 (2013).

97 Shuford, C. M., Hawkridge, A. M., Burnett, J. C., Jr. & Muddiman, D. C. Utilizing spectral counting to quantitatively characterize tandem removal of abundant proteins (TRAP) in human plasma. *Anal Chem* **82**, 10179-10185, doi:10.1021/ac102248d (2010).

98 Tan, S. H., Mohamedali, A., Kapur, A. & Baker, M. S. Ultradepletion of human plasma using chicken antibodies: a proof of concept study. *J Proteome Res* **12**, 2399-2413, doi:10.1021/pr3007182 (2013).

99 Shi, T. *et al.* IgY14 and SuperMix immunoaffinity separations coupled with liquid chromatography-mass spectrometry for human plasma proteomics biomarker discovery. *Methods* **56**, 246-253, doi:10.1016/j.ymeth.2011.09.001 (2012).

100 Qian, W. J. *et al.* Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. *Mol Cell Proteomics* **7**, 1963-1973, doi:10.1074/mcp.M800008-MCP200 (2008).

101 Jones, K. A. *et al.* Immunodepletion plasma proteomics by tripleTOF 5600 and Orbitrap elite/LTQ-Orbitrap Velos/Q exactive mass spectrometers. *J Proteome Res* **12**, 4351-4365, doi:10.1021/pr400307u (2013).

102 Keshishian, H. *et al.* Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol Cell Proteomics*, doi:10.1074/mcp.M114.046813 (2015).

103 Tu, C. *et al.* Depletion of abundant plasma proteins and limitations of plasma proteomics. *J Proteome Res* **9**, 4982-4991, doi:10.1021/pr100646w (2010).

104 Bellei, E. *et al.* High-abundance proteins depletion for serum proteomic analysis: concomitant removal of non-targeted proteins. *Amino acids* **40**, 145-156, doi:10.1007/s00726-010-0628-x (2011).

105 Boersema, P. J., Geiger, T., Wisniewski, J. R. & Mann, M. Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics* **12**, 158-171, doi:10.1074/mcp.M112.023614 (2013).

106 Ostergaard, O. *et al.* Quantitative proteome profiling of normal human circulating microparticles. *J Proteome Res* **11**, 2154-2163, doi:10.1021/pr200901p (2012).

107 Harel, M., Oren-Giladi, P., Kaidar-Person, O., Shaked, Y. & Geiger, T. Proteomics of Microparticles with SILAC Quantification (PROMIS-Quan): A Novel Proteomic Method for Plasma Biomarker Quantification. *Mol Cell Proteomics* **14**, 1127-1136, doi:10.1074/mcp.M114.043364 (2015).

108 Jin, M., Drwal, G., Bourgeois, T., Saltz, J. & Wu, H. M. Distinct proteome features of plasma microparticles. *Proteomics* **5**, 1940-1952, doi:10.1002/pmic.200401057 (2005).

109 Hippius, H. & Neundorfer, G. The discovery of Alzheimer's disease. *Dialogues Clin Neurosci* **5**, 101-108 (2003).

110 Alzheimer, A., Stelzmann, R. A., Schnitzlein, H. N. & Murtagh, F. R. An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde". *Clin Anat* **8**, 429-431, doi:10.1002/ca.980080612 (1995).

111 Tomlinson, B. E., Blessed, G. & Roth, M. Observations on the brains of non-demented old people. *J Neurol Sci* **7**, 331-356 (1968).

112 Tomlinson, B. E., Blessed, G. & Roth, M. Observations on the brains of demented old people. *J Neurol Sci* **11**, 205-242 (1970).

113 Kukull, W. A. *et al.* Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol* **59**, 1737-1746 (2002).

114 Salmon, D. P. & Bondi, M. W. Neuropsychological assessment of dementia. *Annu Rev Psychol* **60**, 257-282, doi:10.1146/annurev.psych.57.102904.190024 (2009).

115 Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412-2414 (1993).

116 Howieson, D. B., Holm, L. A., Kaye, J. A., Oken, B. S. & Howieson, J. Neurologic function in the optimally healthy oldest old. Neuropsychological evaluation. *Neurology* **43**, 1882-1886 (1993).

117 Rubin, E. H. *et al.* A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Arch Neurol* **55**, 395-401 (1998).

118 Masters, C. L. *et al.* Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci U S A* **82**, 4245-4249 (1985).

119 Kang, J. *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733-736, doi:10.1038/325733a0 (1987).

120 Grundke-Iqbal, I. *et al.* Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci U S A* **83**, 4913-4917 (1986).

121 Blennow, K., de Leon, M. J. & Zetterberg, H. Alzheimer's disease. *Lancet* **368**, 387-403, doi:10.1016/S0140-6736(06)69113-7 (2006).

122 Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A. & Evans, D. A. Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch Neurol* **60**, 1119-1122, doi:10.1001/archneur.60.8.1119 (2003).

123 Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-356, doi:10.1126/science.1072994 (2002).

124 Williams, D. R. Tauopathies: classification and clinical update on neurodegenerative diseases associated with microtubule-associated protein tau. *Intern Med J* **36**, 652-660, doi:10.1111/j.1445-5994.2006.01153.x (2006).

125 Min, S. W. *et al.* Critical role of acetylation in tau-mediated neurodegeneration and cognitive deficits. *Nat Med* **21**, 1154-1162, doi:10.1038/nm.3951 (2015).

126 Francis, P. T., Palmer, A. M., Snape, M. & Wilcock, G. K. The cholinergic hypothesis of Alzheimer's disease: a review of progress. *J Neurol Neurosurg Psychiatry* **66**, 137-147 (1999).

127 Markesbery, W. R. Oxidative stress hypothesis in Alzheimer's disease. *Free Radic Biol Med* **23**, 134-147 (1997).

128 Orpiszewski, J., Schormann, N., Kluve-Beckerman, B., Liepnieks, J. J. & Benson, M. D. Protein aging hypothesis of Alzheimer disease. *FASEB J* **14**, 1255-1263 (2000).

129 Yang, H., Lyutvinskiy, Y., Soininen, H. & Zubarev, R. A. Alzheimer's disease and mild cognitive impairment are associated with elevated levels of isoaspartyl residues in blood plasma proteins. *J Alzheimers Dis* **27**, 113-118, doi:10.3233/JAD-2011-110626 (2011).

130 Garcia-Alloza, M. *et al.* Existing plaques and neuritic abnormalities in APP:PS1 mice are not affected by administration of the gamma-secretase inhibitor LY-411575. *Mol Neurodegener* **4**, 19, doi:10.1186/1750-1326-4-19 (2009).

131 Das, P., Murphy, M. P., Younkin, L. H., Younkin, S. G. & Golde, T. E. Reduced effectiveness of Abeta1-42 immunization in APP transgenic mice with significant amyloid deposition. *Neurobiol Aging* **22**, 721-727 (2001).

132 Levites, Y. *et al.* Anti-Abeta42- and anti-Abeta40-specific mAbs attenuate amyloid deposition in an Alzheimer disease mouse model. *J Clin Invest* **116**, 193-201, doi:10.1172/JCI25410 (2006).

133 Ballard, C. *et al.* Alzheimer's disease. *Lancet* **377**, 1019-1031, doi:10.1016/S0140-6736(10)61349-9 (2011).

134 Siemers, E. R. How can we recognize "disease modification" effects? *J Nutr Health Aging* **13**, 341-343 (2009).

135 Julkunen, V. *et al.* Differences in cortical thickness in healthy controls, subjects with mild cognitive impairment, and Alzheimer's disease patients: a longitudinal study. *J Alzheimers Dis* **21**, 1141-1151 (2010).

136 Aarsland, D. *et al.* Frequency and case identification of dementia with Lewy bodies using the revised consensus criteria. *Dement Geriatr Cogn Disord* **26**, 445-452, doi:10.1159/000165917 (2008).

137 Chingin, K., Astorga-Wells, J., Najafabadi, M. P., Lavold, T. & Zubarev, R. A. Separation of Polypeptides by Isoelectric Point Focusing in Electrospray-Friendly Solution Using a Multiple-Junction Capillary Fractionator. *Analytical Chemistry* **84**, 6856-6862, doi:Doi 10.1021/Ac3013016 (2012).

138 Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324, doi:10.1038/nmeth.2834 (2014).

139 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).

140 Ishihama, Y., Rappsilber, J. & Mann, M. Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J Proteome Res* **5**, 988-994, doi:10.1021/pr050385q (2006).

141 Astorga-Wells, J., Landreh, M., Johansson, J., Bergman, T. & Jornvall, H. A membrane cell for on-line hydrogen/deuterium exchange to study protein folding and protein-protein interactions by mass spectrometry. *Mol Cell Proteomics* **10**, M110 006510, doi:10.1074/mcp.M110.006510 (2011).

142 Pirmoradian, M. *et al.* Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol Cell Proteomics* **12**, 3330-3338, doi:10.1074/mcp.O113.028787 (2013).

143 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526, doi:10.1074/mcp.M113.031591 (2014).

144 Trygg, J., Holmes, E. & Lundstedt, T. Chemometrics in metabonomics. *J Proteome Res* **6**, 469-479, doi:10.1021/pr060594q (2007).

145 Huang, T. M., Wu, X. Z. & Pawliszyn, J. Capillary isoelectric focusing without carrier ampholytes. *Analytical Chemistry* **72**, 4758-4761, doi:Doi 10.1021/Ac000599l (2000).

146 Keller, B. O., Sui, J., Young, A. B. & Whittal, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta* **627**, 71-81, doi:10.1016/j.aca.2008.04.043 (2008).

147 Bauer, J. *et al.* Interleukin-6 and alpha-2-macroglobulin indicate an acute-phase state in Alzheimer's disease cortices. *FEBS letters* **285**, 111-114 (1991).

148 Cucullo, L. *et al.* Blood-brain barrier damage induces release of alpha2-macroglobulin. *Mol Cell Proteomics* **2**, 234-241, doi:10.1074/mcp.M200077-MCP200 (2003).

149 Strauss, S. *et al.* Detection of interleukin-6 and alpha 2-macroglobulin immunoreactivity in cortex and hippocampus of Alzheimer's disease patients. *Laboratory investigation; a journal of technical methods and pathology* **66**, 223-230 (1992).

150 Armstrong, P. B. & Quigley, J. P. A role for protease inhibitors in immunity of long-lived animals. *Advances in experimental medicine and biology* **484**, 141-160 (2001).

151 Tucker, H. M., Kihiko-Ehmann, M., Wright, S., Rydel, R. E. & Estus, S. Tissue plasminogen activator requires plasminogen to modulate amyloid-beta neurotoxicity and deposition. *Journal of neurochemistry* **75**, 2172-2177 (2000).

152 Yasuhara, O., Walker, D. G. & McGeer, P. L. Hageman factor and its binding sites are present in senile plaques of Alzheimer's disease. *Brain research* **654**, 234-240 (1994).

153 Bergamaschini, L. *et al.* Activation of the contact system in cerebrospinal fluid of patients with Alzheimer disease. *Alzheimer disease and associated disorders* **12**, 102-108 (1998).

154 Schjeide, B. M. *et al.* The role of clusterin, complement receptor 1, and phosphatidylinositol binding clathrin assembly protein in Alzheimer disease risk and cerebrospinal fluid biomarker levels. *Arch Gen Psychiatry* **68**, 207-213, doi:10.1001/archgenpsychiatry.2010.196 (2011).

155 Thambisetty, M. *et al.* Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Arch Gen Psychiatry* **67**, 739-748, doi:10.1001/archgenpsychiatry.2010.78 (2010).

# Appendix