

Cerebrospinal fluid peptidomics: discovery of endogenous peptides as biomarkers of Alzheimer's disease

Karl Hansson

Department of Psychiatry and Neurochemistry
Institute of Neuroscience and Physiology
Sahlgrenska Academy at the University of
Gothenburg



UNIVERSITY OF GOTHENBURG

Göteborg 2018

Cover illustration by Elin Bergman

Cerebrospinal fluid peptidomics: discovery of endogenous peptides as biomarkers of Alzheimer's disease

© Karl Hansson 2018

karl.hansson@neuro.gu.se

ISBN 978-91-7833-157-4 (Print)

ISBN 978-91-7833-158-1 (PDF - <http://hdl.handle.net/2077/56928>)

Printed in Gothenburg, Sweden 2018

Printed by BrandFactory AB

It's still magic even if you know how it's done

- Terry Pratchett

Cerebrospinal fluid peptidomics: discovery of endogenous peptides as biomarkers of Alzheimer's disease

Karl Hansson

Department of Psychiatry and Neurochemistry
Institute of Neuroscience and Physiology
Sahlgrenska Academy at the University of Gothenburg

Abstract

Neurodegenerative diseases (NDs), the most prominent example of which is Alzheimer's disease (AD), has turned out to be among the greatest challenges for modern medicine. Common characteristics of NDs involve the aggregation of proteins, progressive loss of neuronal cells in specific regions of the central nervous system (CNS) and as a result – cognitive and/or functional decline. Another common feature of most NDs is an extended prodromal stage, in the case of AD believed to be initiated over a decade ahead of noticeable symptoms. Finally; atypical disease presentation and a high frequency in co-morbidities means that specific NDs are generally difficult to define and distinguish. Research would therefore benefit greatly from new biomarkers that can aid in diagnosis, be used for monitoring disease progression, and provide insight into the disease mechanisms. As new disease-modifying therapies are being developed, for example against AD, there will be an increased need for biomarkers that enable earlier and more accurate diagnosis and response to treatment.

Analysis of cerebrospinal fluid (CSF) is valuable to the study of NDs. A multitude of molecules shed by cells are deposited in the CSF, and thus, many processes in the CNS are dynamically reflected in the molecular composition of the CSF. Previous studies have revealed that CSF, besides proteins, contains many endogenous peptides. Being the products of a variety of processes, such as enzymatic protein processing, secretion, and aggregation, these peptides may convey valuable biomarker information.

From an analytical point of view, endogenous peptides are attractive: circumventing proteolytic digestion eliminates a source of analytical variability, and reduces cost and sample preparation time, which are important aspects for establishing assays for clinical research and routine settings. Furthermore, endogenous peptides can readily be isolated from the high-abundant proteins that make up the bulk of the CSF protein contents, by for instance molecular-weight ultrafiltration, thereby allowing a larger volume of CSF peptide extract to be used for LC-MS analysis, improving chances to detect low-abundant peptide species.

The initial aim of this thesis was to develop and optimise methods for isolation, separation, detection and identification of endogenous CSF peptides, with a special focus on low-abundant species. Further, strategies for improved data utilisation and quantitative analysis were also evaluated and subsequently implemented with the goal of identifying endogenous CSF peptide biomarker candidates from clinical cohorts.

Our studies have shown both that the endopeptidome of human CSF is substantially larger than previously indicated and containing a large number of peptides originating from proteins of noted interest in the study of NDs. Further, by means of extensive sample preparation and improved data analysis-techniques we were able to identify a multitude of potential biomarker prospects and, most importantly, three novel biomarker candidates for AD of validated diagnostic value.

More studies are required to further evaluate the identified biomarker prospects for diagnostic value as well as to investigate what their respective presence in CSF may tell about various processes in the CNS. However, the studies included in this thesis have shown that the CSF endopeptidome is a source of information into neurodegeneration with great potential.

Sammanfattning på Svenska

Neurodegenerative sjukdomar, av vilka det mest kända exempel är Alzheimers sjukdom, är bland de största utmaningarna för modern läkevetenskap. Gemensamma karakteristika för gruppen involverar proteinaggregering, progressiv förlust av nervceller i sjukdomsspecifika delar av det centrala nervsystemet och som ett resultat därav – förlust av kognitiv förmåga samt funktionsnedsättning. Ett annat gemensamt kännetecken av de flesta neurodegenerative sjukdomar är en lång pre-symptomatisk period, Alzheimers sjukdom tros till exempel inledas årtionden innan dess att märkbara symptom uppstår. Slutligen, icke-typisk sjukdomsprocess och en hög frekvens av komorbiditet resulterar i att neurodegenerativa sjukdomar generellt är svåra att både diagnosticera och särskilja från varandra. För att förbättra möjligheten att studera den här gruppen sjukdomar behövs nya biomarkörer för diagnos, prognos samt för att kunna observera sjukdomsförloppet och mekanismer involverade i dess utveckling. När nya behandlingar nu börjar dyka upp kommer det även behövas biomarkörer för en tidigare och säkrare diagnos samt för att kontrollera behandlingens effekt.

Cerebrospinalvätska (eller likvor) är ett av de mest använda provmaterialen i studien av neurodegenerativa sjukdomar. En stor mängd molekyler utsöndras till cerebrospinalvätskan från celler i det centrala nervsystemet vilket innebär att processer i hjärnan reflekteras i större eller mindre utsträckning i cerebrospinalvätskan. Det innebär även att man kan använda cerebrospinalvätska för observation av det centrala nervsystemet. Tidigare studier har påvisat att cerebrospinalvätska utöver proteiner också innehåller en stor mängd endogena peptider. De endogena peptiderna tros vara ett resultat av en rad olika processer i det centrala nervsystemet och kan således möjligen vara en potentiell källa till biomarkörer.

Från ett analytiskt perspektiv är endogena peptider intressanta eftersom de tillåter en mindre komplex provpreparation som undviker diverse potentiella felkällor vilka är viktiga aspekter att ta hänsyn till när man försöker utveckla metoder för klinisk rutinanalys. Vidare så kan endogena peptider isoleras direkt från cerebrospinalvätska med hjälp av olika separationsmetoder, som till exempel molekylviktsfiltrering, vilket även medför att man avlägsnar stora delar av proteininnehållet. Genom att på detta sätt exkludera proteiner från provet kan man använda en större volym cerebrospinalvätska, vilket i sin tur innebär att en större mängd endogena peptider ingår i efterföljande masspektrometrisk analys och att lågförekommande peptidspecier således har större chans att detekteras.

Det primära målet i denna avhandling var att utveckla och optimera metoder för att studera endogena peptider i cerebrospinalvätska, med en speciell fokus på lågförekommande specier. Strategier för förbättrat utnyttjande av data och för kvantitativ analys utvecklades och implementerades för ändamålet att identifiera biomarkörkandidater bland de endogena peptiderna från kliniska kohorter.

Våra studier har visat att endopeptidomet i mänsklig cerebrospinalvätska inkluderar ett betydligt större antal peptider än tidigare visats, och även att många proteiner av intresse i forskning om neurodegenerativa sjukdomar representeras av endogena peptider. Vidare så har vi, tack vare de metoder för provpreparation och analysteknik som utvecklats kunnat identifiera en lång rad potentiella biomarkörkandidater. Slutligen, tre biomarkörer för Alzheimers sjukdom har identifierats och utvärderats i kliniska material, med påvisad diagnostisk förmåga.

Fler studier behövs för att utvärdera den kliniska nyttan av de potentiella biomarkörkandidater vi identifierat samt för att undersöka vilka processer i det centrala nervsystemet som resulterar i just dessa peptider. Det går dock att säga från de studier som här har utförts att endogena peptider i cerebrospinalvätska är en källa till information om neurodegenerativa sjukdomar av stor potential.

List of Papers

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- I. **Hansson K**, Skillbäck T, Pernevik E, Kern S, Portelius E, Höglund K, Brinkmalm G, Holmén-Larsson J, Blennow K, Zetterberg H and Gobom J. *Expanding the cerebrospinal fluid endopeptidome*. Proteomics 2017, 17:5.
- II. **Hansson K**, Zetterberg H, Blennow K, and Gobom J. *Turbulent flow chromatography for rapid cerebrospinal fluid sample preparation for clinical peptidomics in Alzheimer's disease*. Manuscript.
- III. Skillbäck T, Mattsson N, **Hansson K**, Mirgorodskaya E, Dahlén R, van der Flier W, Scheltens P, Duits F, Hansson O, Teunissen C, Blennow K, Zetterberg H and Gobom J. *A novel quantification-driven proteomic strategy identifies an endogenous peptide of pleiotrophin as a new biomarker of Alzheimer's disease*. Scientific Reports 2017, 7:1.
- IV. **Hansson K**, Dahlén R, Hansson O, Pernevik E, Paterson R W, Schott J M, Magdalinou N, Zetterberg H, Blennow, K and Gobom J. *The protein-to-peptide ratio improves the performance of tau in CSF as a biomarker of Alzheimer's disease*. Manuscript submitted.

Contents

Abstract	I
Sammanfattning på Svenska	II
List of Papers.....	III
Abbreviations	VII
1. Introduction	1
1.1. Neurodegenerative diseases.....	2
1.1.1. Definition and characteristics	2
1.1.2. Alzheimer’s disease.....	3
1.1.3. Diagnostic challenges	5
1.1.4. Why do we need more biomarkers of NDs	5
1.2. Cerebrospinal fluid	5
1.2.1. Source	6
1.2.2. Sampling.....	6
1.2.3. CSF biomarkers	6
1.2.4. Biomarker candidates	8
1.3. Proteomics	10
1.3.1. Shotgun proteomics	10
1.3.2. Targeted proteomics	11
1.4. Clinical proteomics.....	12
1.5. Peptidomics	13
2. Aim.....	15
2.1. General aims.....	15
2.2. Specific aims	15
3. Methods	17
3.1. Sample pre-treatment	17

3.1.1.	Protein denaturation.....	17
3.2.	Molecular weight cut-off ultrafiltration.....	19
3.3.	Solid phase extraction: principles.....	20
3.4.	Losses and contaminants	21
3.5.	Reversed-phase liquid chromatography	21
3.5.1.	Sample pre-fractionation	22
3.5.2.	Turbulent flow chromatography.....	23
3.6.	Mass spectrometry.....	24
3.6.1.	Ionisation.....	25
3.6.2.	Mass analysis.....	26
3.6.3.	Ion fragmentation methods.....	29
3.6.4.	Modes of operation.....	31
3.6.5.	Hybrid mass spectrometers.....	35
4.	Results and discussion.....	39
4.1.	Paper I: Expanding the CSF peptidome	39
4.2.	Paper II: Turbulent Flow Chromatography	43
4.3.	Paper III: Spectral clustering and pleiotrophin.....	45
4.4.	Paper IV: Tau protein-to-peptide ratio	50
5.	Conclusions	55
5.1.	The CSF endopeptidome	55
5.2.	The peptides	56
6.	Future perspectives.....	57
	Acknowledgements	58
	References	61

Abbreviations

AD	Alzheimer's disease
AGC	Automatic gain control
ALS	Amyotrophic lateral sclerosis
APP	Amyloid precursor protein
A β	Amyloid β
A β ₁₋₄₂	Amyloid β amino acid sequence 1-42
BBB	Blood-brain-barrier
CID	Collision induced dissociation
CJD	Creutzfeldt-Jakob disease
CNS	Central nervous system
CSF	Cerebrospinal fluid
Da	Dalton; g/mol
DDA	Data-dependent acquisition
DLB	Dementia with Lewy bodies
DMS-5	Diagnostic and statistical manual of mental disorders - 5th edition
DTT	Dithiothreitol
ESI	Electrospray ionisation
ETD	Electron transfer dissociation
eV	Electron volt
FA	Formic acid
FDR	False discovery rate
GdnHCl	Guanidinium hydrochloride
HC	Healthy control
HCD	High(er) energy collision induced dissociation
HPLC	High pressure/performance liquid chromatography
IAA	Iodoacetamide
IP	Immunoprecipitation
IWG	International working group
LC	Liquid chromatography
LLOD	Lowest limit of detection
LLOQ	Lowest limit of quantitation
LTQ	Linear ion trap mass spectrometer
m/z	Mass-to-charge-ratio
MALDI	Matrix-assisted laser desorption/ionisation
MAPT / Tau	Microtubule-associated protein tau
MCI	Mild cognitive impairment
MS	Mass spectrometry
ms/ms	Tandem mass spectrometry
MWCO	Molecular weight cut-off (filtration)
ND	Neurodegenerative diseases
NF(L/M/H)	Neurofilament (light/medium/heavy) chain
Ng/RC3	Neurogranin
NIA-AA	US National Institute on Aging-Alzheimer's Association
OT	Orbitrap

PAGE	Polyacrylamide gel electrophoresis
PART	Primary age-related tauopathy
PD	Parkinson's disease
PEG	Polyethylene glycol
PET	Positron emission tomography
PRM	Parallel reaction monitoring
PSM	Peptide sequence match
PSP	Progressive supranuclear palsy
P-tau	Phosphorylated protein tau
PTM	Post translational modification
RF	Radio frequency
ROC	Receiver operating characteristic
RP	Reverse phase
RSLC	Rapid separation liquid chromatography
SDS	Sodium dodecyl sulphate
SIL	Synthetic isotope-labelled (peptide)
SPE	Solid phase extraction
SRM	Single reaction monitoring
TBI	Traumatic brain injury
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TFA	Trifluoroacetic acid
TFC	Turbulent flow chromatography
TMT	Tandem mass tag
T-tau	Total concentration of protein tau
UV	Ultra-violet
YKL-40	Chitinase-3-like protein 1

1. Introduction

As the average human life expectancy keeps increasing globally, disorders correlating with age, such as neurodegeneration, follow suit [1, 2]. The consequence of this relationship is that an expanding part of the population lives longer with one or several conditions, that renders them incapable of normal life and function, causes a great deal of suffering and represents a big expense for society.

The root cause of most neurodegenerative disorders (ND), *e.g.*, Alzheimer's disease (AD), Parkinson's disease (PD) and dementia with Lewy bodies (DLB), are unknown. Diagnosis is commonly impeded by a lack of defined clinical parameters; a deficiency largely stemming from many NDs presenting overlapping pathological manifestations as well as comorbidity, particularly common in AD patients [3]. Thus, there is a need for novel diagnostic tools and biomarkers in the diagnosis, differentiation, study and treatment of neurodegeneration. A precise and early diagnosis would have a great impact on a number of conditions within the realm of neurodegenerative disorders [4-8].

In the case of neurodegeneration, cerebrospinal fluid (CSF) is arguably the preferred diagnostic material. Proximal to the brain, CSF contains a complex mixture of thousands of proteins, peptides, metabolites, salts and a wide range of other components of which approximately 20% of the mass is derived from the central nervous system (CNS) [9, 10], while the remaining 80% comes from the blood. Furthermore, CSF is in a constant flux with a turnover rate that exchanges the entire volume approximately 3-4 times per day, resulting in responses to stimuli in the brain being possible to detect quite rapidly [9, 11]. The usefulness of CSF for the development of novel biomarkers has been demonstrated for AD where a combination of altered CSF concentrations of phosphorylated tau (P-tau), total tau (T-tau) and amyloid- β (A β) are 85-95% sensitive and specific for Alzheimer's disease, both at the prodromal and dementia stages [12]. However, even though there are some candidates under investigation for other NDs [13, 14], so far only AD has been coupled with defined biomarkers. Yet, AD is a heterogeneous disease with a high incidence of atypical presentation and co-morbidities, and the need for both diagnostic and differential diagnostic markers is therefore still great, as with all other NDs.

Since the mid-eighties, progress in mass spectrometry (MS) has led to the technique becoming more and more prominent in the study of complex samples and discovery of novel biomarkers [15, 16]. In particular the development of "softer" analyte ionisation methods, such as electrospray ionisation (ESI) [17, 18] and matrix-assisted laser desorption ionisation (MALDI) has improved the capabilities of studying proteins and peptides [17, 19-21]. From there, the evolution of MS has accelerated to a point where instruments are able to identify thousands of proteins from a single biological sample in the microgram-range are used by many facilities. This, in-turn, has resulted in the possibility of using MS for non-hypothesis driven, or explorative, purposes and thereby introduced (or at least accelerated) the field of proteomics which has revolutionised many areas of biology [22]. In the clinical area however, the progress has been slower; despite almost two decades of proteomic studies of NDs, proteomics has contributed relatively few – if any – new biomarkers to clinical research [3, 13]. This may indicate that proteomic methods and analytical strategies need further development, or simply that a degree of maturation of any field of science is necessary – *i.e.*, researchers need to get the hang of the technology.

The concept of "omics", as in "genomics" and "proteomics", in biology refers to the exhaustive study of a particular set of molecules and their role in a biological process, tissue or whole organism. In proteomics the general aim is to study protein expression, structure and function from observations of various sample materials such as blood, urine, CSF, cell cultures or tissue biopsies [22]. This thesis

another type of omics will primarily be discussed, “peptidomics” or “endopeptidomics”, in which the endogenous peptides present in biological samples are in focus, rather than the proteins from which they derive [23]. There is a discussion on whether peptidomics should be considered a stand-alone subject or be part of the proteomics sphere, a debate that may be regarded purely academic in either case. However, endogenous peptides are ubiquitous in biological systems, sometimes as biologically inert entities, *i.e.*, waste resulting from degradation of proteins, but biologically active peptides are also a vital part of homeostasis [23-25].

In conclusion, the main goal of this thesis was to evaluate the peptidome of CSF as a viable source of biomarkers for NDs. To this end, methods and protocols for efficiently isolating, detecting, identifying and quantifying endogenous CSF peptides in both pooled material and clinical cohorts were developed and employed. The results of all trials performed were a substantial increase of the known constituents of the CSF endopeptidome and importantly, the discovery and evaluation of three potential biomarkers for AD, as well as a large number of not-yet evaluated biomarkers prospects. Furthermore, the methods and protocols themselves; developed for the study of endogenous peptides in CSF but with minor alterations applicable for future studies also in other sample materials.

1.1. Neurodegenerative diseases

Neurodegenerative diseases (NDs) include both dementias and movement disorders. They embody a large and, in many aspects, heterogeneous group of pathophysiological and mental conditions. The DMS-5 (Diagnostic and Statistical Manual of Mental Disorders), published by the American Psychiatric Association, defines dementia as a significant decline in one or several aspects of cognitive performance which negatively impacts daily life and function [26].

Dementia is considered among the top healthcare challenges of today and tomorrow, currently affecting nearly 50 million individuals worldwide and surpassed only by the care and treatment of cancer in cost to society [27, 28]. The number of studies involving the various states of dementia and the amount of funding invested [29] have so far not yielded either a cure, or even any definitive pathophysiological mechanisms. However, novel biomarker candidates are emerging along with an ever-improving ability to study the subtle and intricate mechanisms acting in the human CNS.

1.1.1. Definition and characteristics

NDs are characterised by progressive loss of neurons and/or supporting cellular structures in specific regions of the CNS [26, 30-34]. The causes for this selective cell death are still largely obscure in the case of most NDs. A common feature is the aggregation of proteins or peptides into insoluble bodies. The role of these aggregates is still debated, although the constituents of the aggregates seem somewhat disease-specific and thus suggest an involvement in the pathophysiological process [35-37]. Each particular ND is further characterised by the area of the CNS primarily affected and thereby by the associated symptoms. However; problems in defining and diagnosing specific disorders stem from frequently occurring co-morbidities as well as large variations in pathological and clinical manifestation [38]. Further, atypical presentation of symptoms or symptoms correlating poorly with the extent of pathophysiological progression are relatively common, making neurodegeneration both difficult to diagnose and to study. The latter is an interesting phenomenon suggesting an apparent resistance to dementia in some patients. It has been referred to as the “cognitive reserve” and allow individuals to largely avoid or at least postpone the effects of neurodegeneration [39, 40].

Although there are a few notable exceptions, such as amyotrophic lateral sclerosis (ALS) and Creutzfeldt-Jakob disease (CJD), most forms of ND correlate with age. Above the age of 60, an estimated 5% of the population are diagnosed with AD and the prevalence increases exponentially from then on [41].

1.1.2. Alzheimer's disease

The form of dementia subsequently known as AD was first described in the beginning of the 20th century by Alois Alzheimer as a disorder with progressive memory loss and eventual disruption of executive function affecting the ability to perform normal and basic activities [42, 43]. Alzheimer further described the distinct pathophysiological traits after post-mortem examination of a patient, which today are known as senile plaques, neurofibrillary tangles and extensive cerebral atrophy. However, it was a former colleague of Alzheimer, Emil Kraepelin, who eventually named the disorder after Alzheimer, after the death of the latter in 1915.

Interestingly, in the original article “Über einen eigenartigen, schweren Erkrankungsprozess der Hirnrinde” (roughly: About a peculiar, severe disease of the cerebral cortex), Alzheimer discussed the problem of discerning and discriminating between the disease which came to bear his name and other forms of senile dementia [42]. Co-morbidity is still considered one of the key issues in AD diagnosis and ways to improve its accuracy are in demand for future diagnostics and scientific purposes.

AD is now known to be the most common form of ND. It corresponds to roughly 70% of dementia cases and affect approximately 50 million people worldwide (2016). The disease has a noticeable impact on the gross world product with a direct cost to society comparable to the Swedish national budget [2, 27]. The actual cost is most likely much higher since the disease also directly and indirectly affects individuals with relations to the patient.

Memory and executive dysfunction in individuals above 65 years is the most well known manifestation of AD. However, atypical cases involve distinctly younger individuals, commonly 50-60 years old, and may affect language, vision and/or executive function prior to noticeable impact on memory [44, 45]. Being the most ubiquitous of neurodegenerative disorder resulting in age-related dementia, AD has been extensively studied for nearly 60 years without any breakthrough in treatment (except for symptomatic treatments that do not modify the underlying disease process) or prevention so far [46]. This can partly be attributed to a lack of diagnostic and predictive biomarkers, due to which clinical trials are difficult to conduct. However; because of the improved ability to study complex biological systems, as shown in this thesis, one can hope/expect this issue to be remedied [47].

1.1.2.1. Pathological manifestations and symptoms

The neuropathological hallmarks of AD include extracellular senile plaques mainly consisting of deposits of amyloid beta 1-42 ($A\beta_{1-42}$) and intraneuronal fibrillary tangles of hyperphosphorylated and truncated microtubule-associated protein tau (MAPT or tau) [2, 48-50]. AD pathology presents a complex pattern, the intricacies of which are still largely unknown and debated.

Among the more influential and widely considered/accepted pathological pathways is the so called ‘amyloid cascade hypothesis’ (ACH) [51]. Briefly, the AD pathogenesis scenario in the ACH is suggested to be initiated as a consequence of increased production and aggregation of the $A\beta_{1-42}$ peptide [52]. Initially this increase was believed to be a result of mutations either in the gene coding for amyloid precursor protein (*APP*) itself and/or the genes coding for presenilin 1 (*PSEN1*) and presenilin 2

(*PSEN2*), both involved in the processing of APP [53]. However, although said genetic factors have been recognized as major effectors in early-onset familial AD, it seems as if sporadic AD is less straight forward and that other mutations and genetic risk factors such as the *APOE* $\epsilon 4$ allele [54-56], as well as other mechanisms may be involved [57, 58]. Independent of its derivation, the neurotoxic effects of upregulated $A\beta_{1-42}$ production is believed to result from both small soluble oligomers and from larger depositions (*i.e.*, plaques) [59, 60], albeit that the exact mechanism(s) of toxicity is still poorly understood [61, 62]. The accumulated toxic effects (sometimes referred to as “aggregated stress”) eventually affects other systems, among them neuronal processing and regulation of microtubule-associated protein tau, resulting in aggregation of the same into paired helical fragments and neurofibrillary tangles [62, 63]. The aggregated stress, possibly added to by the presence of tangles, eventually results in neuronal loss, cognitive impairment, dementia and finally death [62]. The ACH has received criticism for not sufficiently explaining some of the mentioned key effectors and mechanisms, but the ACH still manages to provide a causal pathway within and around which further research can be performed to either support or ultimately falsify the hypothesis.

Irrespective of the exact pathogenesis, typical (or sporadic) AD pathology involves loss of neuronal cells, initially in the temporal lobe where hippocampus and entorhinal cortex are primarily impacted [2, 64, 65]. Because of the areas affected, the most noticeable early symptom is impaired episodic memory followed by loss of other cognitive functions such as language comprehension and formulation, which gets steadily more severe as the disease progresses. Progressive loss of neuronal function eventually results in the patient ending up in a completely vegetative state. Death results either from secondary conditions such as respiratory infections (pneumonia) or stroke, or as a direct consequence of dysregulation of life sustaining functions; cause of death in the latter case put down simply as “dementia/senility” [66, 67].

1.1.2.2. Diagnosis

In 2014, the International Working Group (IWG) together with the US National Institute on Ageing-Alzheimer’s Association (NIA-AA) updated their previously defined criteria for the diagnosis of AD [64, 68]. The updated version redefines the concept of dementia but takes into account the previous definition of probable AD based on clinical evaluation focused on cognitive dysfunction, impaired memory and, importantly, the absence of co-morbidities [69, 70]. The main alteration in the updated criteria is the suggestion of introducing neurochemical and neuroimaging diagnostic markers to study two separate biological processes of AD; plaque pathology and neuronal damage [70]. A low concentration of CSF $A\beta_{1-42}$ and a positive signal on amyloid positron emission tomographic (PET) imaging indicates plaque pathology [71-73]. The extent of neuronal damage is indicated by elevated CSF concentrations of total tau protein (T-tau) and hyperphosphorylated tau (P-tau), and reduced fluorine-18 fluorodeoxyglucose ($[F^{18}]FDG$) uptake in the temporoparietal cortex [74]. It should be noted that the criteria involving the employment of diagnostic markers for AD specify use only for research purposes, meaning that implementing said biomarkers in clinical routine analysis is not yet recommended.

The CSF concentrations of the above-mentioned molecules, and the presence indicators measured by neuroimaging techniques, *i.e.*, PET, are considered “core feasible” biomarkers of AD [75]. The term means that these biomarkers are not only well defined for the purpose of diagnosing AD but are also capable of measuring specific aspects of the extent of neurodegenerative progression and may be possible to employ in preclinical stages of the disorder as well [75-77].

1.1.3. Diagnostic challenges

The diagnosis of NDs is generally considered difficult and often relies on the employment of a plethora of complementing diagnostic tools, over a period of time, to reject or confirm a certain disorder [78]. The diagnostic challenges result from the slow and stealthy progression of most NDs and the presentation of atypical symptoms [13]. This is often exacerbated by variation in disease definition, which makes differentiation of disorders difficult.

Several NDs, *e.g.*, AD, frequently present co-morbidity in pathological manifestation; for instance, AD often develops alongside other sources for cognitive decline such as dementia with Lewy bodies (DLB) and vascular dementia [13]. The ability to identify individuals likely to develop neuropathology in later life is key for successfully studying this class of diseases [64]. A number of genetic variations that correlates with an increased risk of developing an ND have been identified [56, 79-81]. Unfortunately, the current diagnostic markers for AD (the only established biomarkers for any ND) do not correlate well with the earliest clinical and cognitive symptoms of the disease, known as mild cognitive impairment (MCI) [64, 82].

Atypical presentation of AD is another challenge because it may involve the presence of the classical plaque-tangle pathology, but cognitive functions remain largely intact. There are also subtypes of the disease which are still considered AD, but that are predominantly affecting other areas – such as primary age-related tauopathy (PART), which mainly affects the limbic system leaving hippocampus relatively unaffected [83].

1.1.4. Why do we need more biomarkers of NDs

In theory, a biomarker can be any characteristic, that may be employed to assess the state of a biological system. Biomarkers are used in diagnosis, prognosis, monitoring of disease progression and target engagement for drug treatment in a clinical setting, or to characterise and explore biological systems for primarily scientific reasons [84].

Due to the progressive and often slow disease development combined with a high frequency of co-morbidities, atypical presentations and individual resistance results in NDs in general and AD in particular being hard to diagnose accurately. The clinical diagnosis can still only be confirmed after post-mortem examination of the patient's brain [85].

There are currently few biomarker candidates for NDs; only for AD have reliable biomarkers been established [12], making the selection of individuals for clinical trials both difficult and error prone. Cohorts commonly contain a number of false positives and negatives, which hampers the study of this category of disorders [13, 86].

The current hypothesis is that the majority of NDs are initiated long before the first noticeable symptoms– meaning that it is only towards the later stages that the disease is diagnosed, with confirmation commonly achieved until post-mortem examination. However, even though a number of genetic risk factors have been identified [79, 87-89], there is currently no reliable way of telling whether someone will develop an ND later in life.

1.2. Cerebrospinal fluid

Examination of cerebrospinal fluid is employed as a pathway for observation of chemical changes within the CNS. CSF is the recipient of a large number of molecules shed by cells in the CNS and is therefore

considered to at least partially reflect many of its processes [10, 90]. However; for MS-based proteomic analysis, CSF also presents an analytical challenge; the concentrations of protein/peptide species in the CSF span between eight and nine orders of magnitude and it is reasonable to assume that the diagnostically relevant markers are present in the lower end of this range [91]. Compared to for instance an immunoassay, in which the target proteins are selectively isolated and detected, explorative clinical proteomics is designed to detect the entire protein complement of a sample and to subsequently attempt to determine each peptide's diagnostic value.

1.2.1. Source

It is hypothesized that the bulk of CSF is produced in the choroid plexus and that interstitial fluid from the surrounding nervous tissue contributes by draining into the CSF [9, 92]. Besides acting as a transport pathway for various CNS products, as CSF surrounds the entire brain and fills grooves and cavities, the fluid also functions as a stabilising medium and offers cushioning support in the event of impact or other forms of trauma. CSF is considered an “ultrafiltrate of plasma” since it is largely derived from fluid passing over the blood brain barrier (BBB). However, it also contains a large amount of products and waste from the CNS, *i.e.*, proteins, peptides, salts, neurotransmitters and cell debris, which is produced locally [90, 93, 94]. With a high turnover-rate of the latter (~500 mL per day in healthy adults) both major and minor events in the CNS are soon reflected in, and can be monitored through, the CSF [95, 96]. Traumatic brain injury (TBI) such as being knocked out in boxing shows up in CSF as an increase of various intracellular neuronal proteins (*e.g.*, tau) over a period of a few weeks

1.2.2. Sampling

Cerebrospinal fluid is extracted by means of lumbar puncture, which is a safe procedure with post-lumbar puncture headache as the only potential complication occurring in 2-20% of patients (varying rates across studies) [97], but the only study done in a blinded fashion (comparing actual with sham lumbar punctures) found no difference in rates of headache [98]. Twelve mL of CSF per patient was extracted via a needle inserted into the dural space between lumbar vertebrae L3/L4 or L4/L5 according to a standardised procedure. The CSF was collected in polypropylene tubes, insoluble material was subsequently removed by centrifugation at 2000 g and 4 °C for 10 min followed by transfer and aliquoting of the supernatant into new tubes. The CSF was finally stored at -80 °C.

1.2.3. CSF biomarkers

Although there is currently no diagnostic biomarkers for any ND fully implemented in clinical routine analysis, the biomarkers for AD described in this chapter have been extensively studied and characterised for this research purpose [12, 77, 86]. The question is hence ‘when’ rather than ‘if’ these neurochemical markers will be employed, or at least accepted, globally in a clinical setting.

1.2.3.1. Amyloid β peptide 1-42

A β ₁₋₄₂ is a pathophysiological biomarker for amyloid pathology measured in CSF and is used for indicating the presence and extent of senile plaque formation in the CNS. The A β ₁₋₄₂ peptide is the result of sequential proteolytic processing of the transmembrane amyloid precursor protein (APP) by β -secretase (BACE) followed by γ -secretase, referred to as the amyloidogenic pathway [99]. In a study published in 2011 by Portelius *et al.* it was shown that CSF contains a large number of APP fragments, indicating that the processing of APP is complex and that several mechanisms for its proteolysis are at work [100]. The APP family of proteins have been extensively studied (primarily in various mouse models) over the past decades [101]. The proteins are highly expressed in neurons and are upregulated

during neuronal differentiation and maturation [102]. Other than that the exact function of APP is unknown. There is little evidence of APP itself being involved in AD pathological process. It is rather the dysfunctional (pathological) processing that results in increased concentrations of $A\beta_{1-42}$ correlating with AD pathology.

Studies have shown that, in individuals with atypically high or low total CSF concentration of A β peptides, $A\beta_{1-42}$ may better correlate with the total A β peptide concentration and therefore result in a mis-diagnosis [103, 104]. A β consisting of 40 amino acids ($A\beta_{1-40}$) is the most abundant A β peptide in CSF and is generally believed to be non-pathological. By employing the $A\beta_{1-42}/A\beta_{1-40}$ -ratio it is possible to compensate for variations in A β peptide concentration levels resulting from processing of APP and this may potentially lead to a more precise diagnosis [103, 104]

1.2.3.2. Total tau (T-tau)

Microtubule-associated protein tau has a range of functions for assembly, as transient stabiliser and as mediator of transport functions, all connected to the intracellular microtubule network [105]. The protein is predominantly expressed in neurons, in a substantially higher degree in the non-myelinated cells of the neocortex (grey matter) as compared to those in the cerebellum and myelinated cells of white matter [106, 107]. The function of the protein is associated with its degree of phosphorylation (see the next section) but it has been suggested that tau is also dependent on its ability to oligomerise in order to form essential interactions [108, 109].

An accumulating body of literature indicates the involvement of specific tau forms and modifications in neuronal dysfunction and pathologies [110]. A key study, in which large volumes of CSF were fractionated by high pressure/performance liquid chromatography (HPLC) and analysed by western blot using different antibodies, revealed the presence of a multitude of N-terminal and mid-domain tau fragments [111]. Several of these fragments appeared to differ in abundance between AD patients and controls, indicating that the tau molecular heterogeneity, reflected in the CSF may contain important biomarker information.

The total CSF concentration of tau protein (or protein fragments containing the respective antibody epitopes) referred to as “T-tau” is a marker of neuron axonal damage [112-116]. This means that T-tau is not specific to AD or even neurodegeneration, but correlates well with the extent of brain injury and neurological damage [77, 117]. The protein concentration in CSF increases when the microtubule-transport system and neuronal cell membrane is disrupted, *e.g.* due to disease or trauma. In AD, T-tau is therefore employed to measure the intensity of neuronal degradation. However, considerably elevated levels of T-tau can also be seen as a consequence of more rapidly progressive NDs such as CJD [118] and in acute traumatic brain injury [117, 119]. Hence, T-tau needs to be complemented by other markers to allow for a more specific diagnosis.

1.2.3.3. Phosphorylated tau (P-tau)

The function of tau is to modulate microtubule assembly, to enable microtubule-based axonal transport in neurons and to be one of the means of cellular structural dynamics [120-122]. The functionality of tau for these purposes is dependent on both stable and transient phosphorylation of specific sites by one or several protein kinases [122], which reduces tau’s microtubule-binding ability (or affinity) along with its ability to stabilise the polymers, allowing for dismantling and restructuring of the microtubule network [123, 124]. An interesting model for tau’s cellular function, both as a stabiliser of microtubules and mediator in axonal transport, but also for other forms of signal transduction [125], is a behaviour

known as “kiss-and-hop” [126], where the interaction between tau and particular microtubule is around 40 μ s long after which tau moves on to another interaction.

A prime candidate for the role as tau “phosphorylator” is glycogen synthase kinase 3 β (GSK3 β), which has been shown to both co-localise with the protein [127, 128] and to employ tau as a substrate [129-131]. Overexpression of GSK3 β in mouse models correlates with an increase in the phosphorylation in a number of sites, possibly indicating the involvement of GSK3 β , or another protein kinase, in the pathologic phosphorylation of the tau protein [122].

Tau phosphorylation is key to both the function and dysfunction of the protein [132]. The presence of truncated and hyperphosphorylated tau in neurofibrillary tangles (NFTs) suggests that at some point, for some reason, the regulation of tau phosphorylation becomes dysfunctional, resulting in an inability to bind to microtubules and to carry out its normal tasks [133]. It is unknown whether the NFT aggregates are the result of spontaneous formation (possibly due to the high degree of phosphorylation or as a result of failed clearance [51, 134]), whether they are actively formed as part of an immune system response, or whether there is a completely different mechanism [116, 133-136]. Either way, an increase in the CSF concentration of P-tau correlates with AD pathology [137], and is for that reason employed as a biomarker for the same.

1.2.4. Biomarker candidates

The section below describes a number of promising biomarker candidates currently under investigation for their potential applicability for the diagnosis of neurodegeneration.

1.2.4.1. Neurofilaments

The neurofilaments (NFs) are a type of intermediate filament and one of the triad of polymeric filaments (along with microtubules and microfilaments) that constitute the so called cytoskeleton¹ of neuronal cells [139]. Differing from the other two filaments, NFs are exclusively expressed in neurons and are especially well represented in large myelinated ditto [140], where it is the most prominent cytoskeletal protein [31].

There are three subunits of NFs suitably named in order of ascending molecular weight; neurofilament light (NFL), medium (NFM) and heavy (NFH) chain. The initial expression of each subunit seems to correlate with the time-dependent requirements of the embryonic neuritic development [141]. The assembly of the NFs is still not well understood but it is believed that the first step involves dimerization of NFL followed by coupled polymerisation with either NFM or NFH resulting in a heteropolymer with structural properties depending on the ratio of NFL to NFM/NFH [142]. Neurofilaments have been reported *not* to assemble into proper filaments *in vitro*; a drawback since it does not allow to study NFs in tissue culture thereby hampering biomarker development [142]

Neurofilament light concentration, similarly to T-tau, increases in CSF as a result of extensive axonal damage and can be employed as a complement to the same in diagnosis of AD [31]. Cerebrospinal fluid concentrations of NFL correlate well with AD pathology, increasing stepwise from healthy control (HC)

¹ “Cytoskeleton” is an umbrella term for a number of different proteins which are to some extent tasked to function as a cellular scaffolding, giving structural support and rigidity to the cell. The name is somewhat misleading since the cells “skeleton” is also crucial in cellular transport and is, compared to its vertebrate namesake, a highly malleable and dynamic construct [138] Shay, J., *the Cytoskeleton*, Springer Science & Business Media 2013, [138] *ibid.*.

to MCI and from MCI to AD [86], however with some overlap between groups. What is most interesting is that NFL is detectable in plasma and further, is able to separate groups to a similar degree as in CSF [12, 143, 144]. However; NFL is not a good marker for differential diagnosis; NFL concentration in CSF can satisfactorily separate HCs and AD pathology-positive patients but does not allow for distinguishing between HCs and tau-positive patients, nor between AD and other neurodegenerative dementias [114].

1.2.4.2. Neurogranin

The calmodulin-binding protein neurogranin (Ng) is involved in long term potentiation/depression of dendritic spines by modulating Ca^{2+} concentrations in the synaptic cleft [145, 146]. Further, Ng is believed to affect neuronal plasticity and thereby cognitive function, but the mechanism by which Ng carries out this role is not fully understood [146].

Synaptic dysfunction and depletion are pathophysiological traits of AD, and have been shown to be among the early steps of AD pathogenesis [54, 147-149]. The severity of cognitive impairment, as well as amyloid pathology have been shown in several studies to correlate well with the CSF concentrations of Ng [30, 150-153]. This suggests that biomarkers correlating with synaptic loss, such as Ng, could also be employed to indicate early onset of AD as well as, to some degree, severity and rate of decline of cognitive impairment [154, 155].

Further, according to Wellington *et al.* (2018), the CSF concentration of the protein is elevated in typical AD compared to atypical cases ($P=0.004$), and is specific for AD compared to other NDs suggesting that Ng can be employed for differential diagnosis as well as for determining typical or atypical AD [156, 157].

1.2.4.3. Chitinase-3-like protein-1

Chitinases constitute a group of glycoside hydrolases with the primary function of degrading chitin² [159]. In mammals the chitinases are expressed in various cell types, most likely as a defence against parasitic insects, fungi and helminths by challenging them when present in the body and to reduce the inflammatory response and allergic reaction induced [160]. In addition to chitinases, mammals also express a group of enzymatically inactive, but structurally similar chitinase-like proteins (CLPs). Their exact biological function is unknown, but evidence suggests that they are involved in immunoregulation [161, 162]. Chitinase-3-like protein-1, also known as YKL-40, is significantly upregulated (in what source depending on specific condition) in patients with asthma/allergy, infections from certain bacteria, arthritis and various fibrotic-disease [161, 163] as well as in a number of inflammatory conditions and cancer [164-166].

YKL-40 has previously been suggested as a marker for TBI and multiple sclerosis [165, 167] but is now being evaluated as a novel CSF biomarker for other types of neurodegeneration, including AD [168-170]. Tests indicate that, although YKL-40 primarily increases with age, it also correlates with AD, FTLT, and vascular dementia [171], importantly without a similar correlation with PD [170, 172]. Thus YKL-40 is potentially useful for differential diagnosis as well as indicating that AD and PD cause different neuroinflammatory responses.

² Chitin is the main constituent of the exoskeleton and cell wall of arthropods, nematodes and fungi. Chitin is one of the most commonly occurring natural polymers on Earth, surpassed in abundance only by cellulose [158] Tharanathan, R. N., Kittur, F. S., Chitin—the undisputed biomolecule of great potential. 2003, [158] *ibid*.

1.3. Proteomics

In order to overcome the various issues, old and new, involved in studying biological sample materials the field of proteomics is continuously introducing new analytical strategies or is updating older ones. Favourite techniques emerge and fall out of favour in the scope of a few years, just to be re-invented or simply return to fashion and make a comeback. The analytical approaches employed in the work performed in this thesis are described in detail in other sections, but a brief introduction to the main proteomics strategies is included in the section below.

1.3.1. Shotgun proteomics

In shotgun proteomics the building blocks of proteins are studied to gain information on the larger structure, a concept known as “bottom-up proteomics”. Initially the whole protein content of a biological sample material is degraded with proteolytic enzymes into peptides, which are subsequently separated over an HPLC column and analysed by MS. Following analysis, the data is most commonly processed by a proteomics software. The software translates the analytical information (*e.g.* intensity, mass and charge of detected ions) from individual *ms/ms*-spectra to peptide-sequences [173]. The peptides, in turn, infer information on the proteins originally present in the sample – that is, the presence of a certain set of peptides (or just a single unique peptide sequence) detected by the MS indicates that the sample contained a particular protein [22, 174, 175].

The main reason for taking this seemingly roundabout route to study proteins is that the typical full-length protein is quite a massive molecular structure; difficult to separate, ionise and fragment efficiently. Relatively novel techniques do allow for top-down proteomics by MS; however, extensive sample preparation and separation as well as MS method optimisation requirements makes for low throughput [175]. In comparison, degrading the proteins into bite-sized peptides ensures relatively straight-forward preparation, standardised HPLC separation and that a large section of the analytes is ionised and available for MS analysis.

The shotgun approach does however require a considerable amount of already available information. This is because most proteomics software carrying out peptide identification employ peptide libraries for comparison with analytical data [174]. The libraries are created by performing *in silico* degradation of a known proteome (*e.g.* the human CSF proteome). Virtual peptides are generated through simulating protein cleavage corresponding to that resulting from the specific proteolytic enzyme employed for degrading the proteins in the actual sample. Databases containing the proteomes of whole organisms have been generated from genomic data. However; mutations, post translational modifications (PTMs) and splice variations requires these databases to be continuously updated and verified [176, 177].

A number of techniques allowing for peptide and protein quantitation in shotgun proteomics have been introduced over the last two decades making proteomics much more useful in clinical studies [178-180]. MS is not, as previously mentioned, an instrument that allows for quantitation unaided – *i.e.*, by directly comparing signal output between different peptides within a sample or the same peptides between samples. The reason for this lack is the high variability stemming from various individual steps along the analytical path.

Two main approaches to overcome this problem have crystallised: “Labelling-based” and “Label-free” quantitation. Labelling-based quantitation employs incorporation of heavy stable isotopes - either through *in vivo* metabolic integration by adding labelled amino acids to the diet, or chemically by labelling peptides *in vitro* during sample preparation [180]. The peptides in the labelled sample can then

be compared to their counterparts in an unlabelled or differently labelled sample and signal intensity can be used for relative quantitation (a step that requires extensive data processing by specialised, and expensive, software) [181]. In contrast the label-free approach relies on peptide peak area/height during chromatographic separation and spectral counting following MS/MS analysis for comparing analytes from different samples [182].

All in all, shotgun proteomics has become the approach of choice for large scale proteomics studies due to its high throughput, ease of preparation and broad analytical range. Methods for relative quantitation has improved the usefulness of shotgun proteomics for studying protein/peptide expression and turnover, and made biomarker discovery a substantially simpler task.

1.3.2. Targeted proteomics

The previously described proteomics methods are primarily aimed at discovery-based studies and meant for allowing identification of large numbers of peptides and proteins from complex samples by employing elongated and multi-dimensional separation protocols. However, despite the introduction of several techniques for improving quantitative abilities of MS, absolute quantitation (*i.e.*, exact determination of the concentration) still requires a fundamentally different approach, and may still not be absolutely achievable³ [183].

The current methods for relative quantitation do not allow for determining exact concentration of analytes, but only the difference in quantity between samples. Targeted proteomics, as the name suggests, is the strategy employed when monitoring a set of pre-determined analytes, usually a comparatively small number relative to the whole proteome/peptidome of the sample material. Compared to other proteomics strategies, the point of targeted proteomics is to gather as much information as required to confirm the presence or to allow for concentration determination of the analytes [24, 184]. The latter is commonly made possible by introducing a standard of known quantity at some stage during sample preparation. The standard contains one or several amino acids carrying heavy isotopes, most commonly of carbon, nitrogen or oxygen. The result is a peptide identical in chemical properties but differing slightly in mass, meaning that it will co-elute, and most importantly, ionise at the same time and (ideally) to the same extent as the native peptide [185-187]. In essence; the two peptides are detected simultaneously but will be distinguishable by the mass difference. Hence, if the concentration of the standard is known the concentration of the native peptide can be determined by comparison [24, 183, 188, 189].

Except for the addition of known quantities of an internal standard, the main difference is usually, as hinted, in the setting of the MS. The first MS-protocol step commonly involves ensuring simultaneous detection of the precursors of the sought after native peptide and its corresponding synthetic standard over their entire elution period [190]. Then, depending on the instrument, one or several fragments of each peptide are detected and subsequently employed for quantitation by comparison, usually of intensity over the period of detection (referred to as total peak area) [191-193]. The main difference lies

³ The question of whether absolute quantitation of complicated analytes using mass spectrometry is at all achievable is not going to be discussed here; it involves elongated discussions on chromatographic behavioural alterations resulting from differences in isotope-content and point-to-point (on milli-second scale) measurement variations of analyte versus internal standard concentrations during ionisation and mass analysis. Albeit extremely important, and enjoyable, to discuss the concept further in order to finally solve the problem, it is possible to consider the currently employed method for “absolute quantitation” sufficiently accurate for its purpose, *e.g.*, to monitor concentration variations of peptides/proteins between samples to determine possible disease involvement.

in whether or not a single product ion is sufficient to identify the analyte or if several fragments are required.

Except for the possibility of quantifying analytes, the main advantage is that the targeted approach usually requires short liquid chromatography (LC)-gradients and thus allows for high sample throughput. Depending on the instrument and the requirements to successfully identify the analytes, an analysis round may take only a few minutes and still allow for highly accurate and sensitive quantitation of the targeted analytes, providing there are not too many.

1.4. Clinical proteomics

The field of clinical proteomics is encumbered with the theoretically straightforward task of determining which proteins are expressed, to what extent, and how they are subsequently modified in healthy contra non-healthy individuals [194]. In essence, this means monitoring the total protein complement from cells, tissues, organs and whole organisms and even further, whole populations and population sub-sets, such as forgetful contra non-forgetful grandmothers [194]. This may yield both insights into disease mechanisms and lead to the development of tools for improved diagnosis, *i.e.*, protein biomarkers and in the end to treatments and cures. Although it might seem an insurmountable enterprise the fact is that great progress has already been made.

The field has advanced considerably with increased knowledge of biological systems to the development of ever-improving technical solutions which forms a positive feed-forward loop [195-197]. For instance, the enormous undertaking of the Human Genome Project impacted nearly every branch of biology as well as other sciences, a range of technologies were improved or developed and the collective understanding of information flow within living organisms was greatly enhanced [198, 199]. However, one of the greatest lessons was that the genome of an organisms says surprisingly little about its phenotype; that is, its actual protein expression [196]. An example of this increased divide between genome and proteome is found in nomenclature, where the original meaning of proteomics was “the set of proteins encoded by the genome” [200]. For this reason, a deal of focus has shifted towards improving the ability to study the proteome, employing genetic information to aid in protein identification, which in turn allows for an increasing understanding of gene-expression.

The construction of databases for proteomics studies has largely been dependent on genetic information, to find out what proteins are expressed in a particular organism, tissue or cell type. [22, 194]. However, phenotypic information not inferable by solely studying the genome, or at least not directly inferable from the information derived from genomics (with our current capacity), *e.g.* post translational modifications, intra- and inter-protein interactions, enzymatic cleavage of proteins into bioactive peptides etc., must be derived through proteomic studies. This means that the proteome contains information on the functionality of an organism that currently stands separate from genomics [196].

In clinical proteomics the focus is aimed at studying protein involvement in health and disease; discovering the difference in protein expression, modification and behaviour induced in individuals by a particular condition [194]. The purpose is to identify pathophysiological pathways, which in practice means that a particular sample material, believed to be closely involved in the disease process (or to reflect the process by proxy) is studied in detail and statistical analysis is perpetrated.

1.5. Peptidomics

As with proteomics (and all other “omics”), the aim of endopeptidomics is to reach complete understanding of the “whole point” of the particular set of analytes, *i.e.*, the endogenous peptides in a given organism. The fluxes in concentration depending on the state of the organism (age, health, sex, etc.) as well as their specific functions and interactions are all concepts sought to be understood [201-203].

Similar to the connection between genome and proteome, the link between proteome and peptidome is seemingly quite clear; that is, one results from the other. Endogenous peptides *do* result from proteins through proteolytic degradation either specific or “random”, depending on the purpose of the product. However, just as with genomics and proteomics, the actual process is much more complex than this explanation lets on [23, 202, 204].

The endogenous peptide content, the endopeptidome, of CSF originates to an unknown degree from processes in and around the CNS, but also from other parts of the body, as large portions of the CSF content is a filtrate passed over the BBB [93]. In general, a large number of intra- and extracellular processes are involved in protein degradation for various purposes, being it simply for removal of “waste proteins/peptides” or recycling of amino acids, for generating bioactive peptides or for various immune-system functions [205, 206]. The remains of such ongoing processes in the CNS can be detected in CSF and are, in some cases, possible to use as biomarkers to determine the state of the machinery upstream, the most prominent example being $A\beta_{1-42}$ or P-tau 175-190 [93] (Hansson *et al.*, manuscript submitted).

Finally, the fact that there is a large number of endogenous peptides present in CSF [207], which functions are largely unknown, can be a complicating phenomenon in proteomics. Large numbers of peptides such as hormones, neurotransmitters and neuromodulators are bioactive in themselves and degrading them through proteolysis and simply employing the remainder to infer information on the originally expressed protein is to lose information on the studied system [23]. Proteomics and peptidomics are compatible areas of study and should be combined to generate the maximum amount of information.

2. Aim

2.1. General aims

The aim with this thesis is to investigate the endogenous peptides present in human cerebrospinal fluid, the CSF endopeptidome, and to evaluate their usefulness in the diagnosis of neurodegenerative diseases, primarily Alzheimer's disease.

2.2. Specific aims

- I. Develop and/or optimise methods for peptide isolation, peptide separation and peptidome deconvolution, mass spectrometric analysis and identification of endogenous peptides from raw ms/ms-data.
- II. Identify and evaluate endogenous peptide biomarker candidates through targeted mass spectrometric analysis
- III. Develop online liquid chromatography-based sample preparation protocols for rapid isolation of endogenous peptide biomarkers for robust clinical routine analysis by mass spectrometry

3. Methods

The main body of work performed in the studies included in this thesis aimed at developing and/or optimising methods and protocols for various purposes in the study of endogenous peptides in human CSF. The protocol for discovery peptidomics here developed and employed is superficially similar to a standard shotgun proteomics protocol. There are however fundamental differences; in peptidomics, as compared to proteomics, analytes (peptides) are not generated but acquired since they are already present in the sample, and most importantly analytes do not directly infer information in a hierarchical manner [23]. This means that endogenous peptides are primarily studied as stand-alone entities, which can potentially be employed to infer information on the system in an indirect manner.

3.1. Sample pre-treatment

Biological samples nearly always constitute a complex mixture of components wherein the analytes of interest make up a smaller or larger fraction of the whole. Once the sample has been extracted, it is therefore commonplace to perform some degree of chemical or mechanical treatment in order to either remove components whose presence would otherwise be detrimental to analysis and/or to make the sought-after analytes more readily available for analysis

The mass spectrometric instrument employed in most of the work is sensitive both in the sense of being able to detect low-abundant analytes but also in being easily interfered with. Hence, a large section of the components of the original CSF sample had to be degraded and/or removed to improve analysis quality and robustness, or simply to make sure the HPLC and/or MS would not clog or break down.

3.1.1. Protein denaturation

A protocol for selective acquisition of endogenous peptides developed by Mikko Hölttä *et al.* (2012) was adapted to some degree from a standard proteomics protocol. This means that the protocol included both protein aggregate-disruption through employment of a chaotrope as well as reduction and alkylation of bridges to avoid spontaneous folding and forming of novel protein aggregates – meant to prepare the sample proteins for proteolytic degradation [208]. Over the course of the studies carried out here, the protocol was altered to some extent, primarily to accommodate a larger volume of CSF per sample but also to re-adjust the active concentration of reagents used for protein denaturation.

The chemical disruption of protein aggregates employed in the protocols used here is performed in two steps; the first is meant to break up protein quaternary, tertiary and, to some degree, also secondary structures. To this end is a detergent or chaotropic agent (chaotrope), for example sodium dodecyl sulphate (SDS), urea or various guanidinium salts commonly employed; here was primarily guanidinium hydrochloride (GdnHCl) used [209]. The chaotrope acts through altering protein-protein and/or protein-water interaction by causing either water molecules or the proteins to rearrange around the chaotrope, and making non-covalent interactions (mainly van der Waals interactions) between proteins less energetically favourable [209-211]. Employing a sufficiently high concentration of chaotrope thereby results in both aggregate disruption and some degree of protein unfolding – leaving only covalent interactions [209, 212].

The second protein structure-disruption step is meant to unfold the proteins completely and to permanently keep them from re-folding by breaking cysteine-cysteine di-sulphide bridges [213]. Cysteine residues present a reactive sulfhydryl group which readily oxidises other sulfhydryl group resulting in intra- and inter-protein covalent disulphide bridges between cysteines [214]. The disulphide

bridges need to be broken to allow for protein unfolding. However, since the formation of di-sulphide bridges are energetically favourable the sulfhydryl groups also need to be made permanently non-reactive in order to avoid spontaneous formation of new bridges. As usual there are many ways to approach the issue, but in general disulphide bridges are initially broken through the action of a reducing agent [213, 215-217] (see figure 1), such as dithiothreitol (DTT), 2-mercaptoethanol (BME) or tris (2-carboxyethyl) phosphine hydrochloride (TCEP) – just to mention a few.

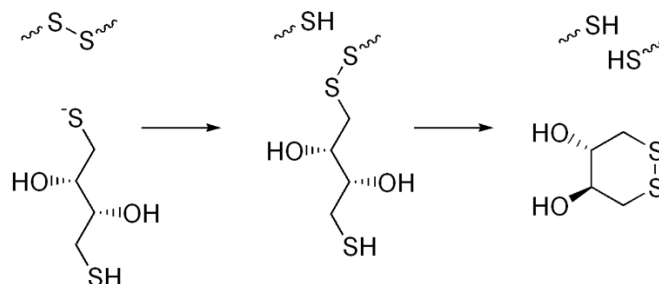


Figure 1: Reduction of an arbitrary disulphide linkage through two-step thiol-disulphide exchange induced by DTT

The sulfhydryls are subsequently rendered non-reactive (“are capped”), commonly through covalent binding of an alkyl group to the free sulfhydryl group; a process known as alkylation [218, 219]. In this case the alkylating reagent was iodoacetamide (IAA), which causes the non-reversible addition of a carbamidomethyl group to the cysteine sulphur residue under alkaline conditions (see figure 2).

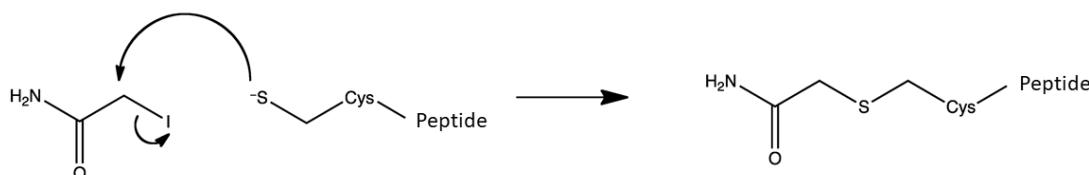


Figure 2: Carbamidomethylation of a cysteine sulfhydryl group at alkaline pH with IAA

The expected/desired effect of these treatment-steps is more or less to completely unfolded proteins. The side effects involve the alteration of the expected mass of each cysteine during peptide identification. The alteration of mass corresponding to carbamidomethylation is an addition of approximately 58 Da to cysteines, which must be included as a static modification or result in erroneous peptide identification.

In *proteomics* these preparation steps are performed to improve the enzymatic digestion of proteins in to peptides prior to analysis. By unfolding the proteins, more cleavage sites are made available and a more complete digestion is achieved. In *peptidomics* on the other hand the hypothesis was that the same treatment would result in less endogenous peptides being tied up in aggregates and allow for better peptide recovery after using a molecular weight cut-off (MWCO) filter. However, trials in which reduction and alkylation were left out of the protocol showed little effect on the number of identified endogenous peptides (data not shown). Since the trials were non-quantitative it was not possible to tell if avoiding these steps actually caused losses, general or in specific peptide species, so the practice of reducing and alkylating was mostly retained. In contrast, leaving out aggregate disruption by means of chaotropic treatment did impact identification negatively. A possible clue to the cause of this was

observed during an experiment with turbulent flow chromatography (see below) where the non-treated samples caused immediate blocking of the separation column.

3.2. Molecular weight cut-off ultrafiltration

In terms of complex biological constituents, CSF primarily contains a small number of highly abundant proteins, *e.g.* albumin and immunoglobulins, as well as somewhere between 2000-3000 less abundant protein species [9, 10, 91, 220-224]. The endogenous peptides we sought to study most likely constitute only a small, albeit highly diverse, fraction of the total mass [90, 207, 225-228]. Since a peptidomics protocol seeks to isolate the peptides, one of the first steps is to separate the small fraction of peptides from the vast bulk of proteins.

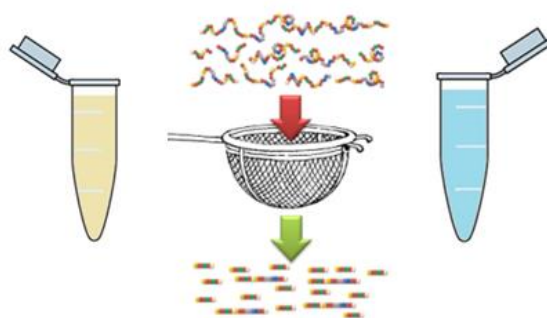


Figure 3: Visualisation of the principles of molecular weight cut-off filtration, accurate to a given value of “accurate”.

Several techniques have been applied for the purpose of selectively isolating endogenous peptides from proteins, or rather, to isolate peptides and/or proteins below a certain mass-range from those above (there is of course a span where the masses of large peptides and small proteins overlap, but intricacies of this type are beyond the scope of this thesis). The primary approaches for isolating *small* endogenous peptides are protein depletion, size exclusion chromatography (SEC) [229] and MWCO ultrafiltration [24]. There have also been relatively recent reports of protocols for separating peptides as small as 500 Da using gel electrophoresis techniques [230] which might be of future interest.

There are advantages and disadvantages associated with each method but a limited amount of time to investigate and evaluate each one. Eventually, we chose to primarily focus on and optimise an MWCO-based separation protocol. The choice was made partly because of experience from earlier *in house* studies by Mikko Hölttä and colleagues, where the efficacy of various MWCO filters for the purpose of peptide separation was tested and subsequently employed [226, 228, 231, 232]. Further, there is a relatively large amount of data supporting the functionality of MWCO for peptide isolation in complex biological samples [222, 233-237]. Finally, the MWCO filters are (typically) single-use cartridges, available in a range of filter sizes and sample loading volumes, which allows for straightforward up- or down-scaling and general ease-of-use. The Amicon® ultracentrifugal filters we eventually selected employed a membrane of regenerated cellulose, and a collection tube of polypropylene, making for a relatively inert system with small analyte losses and ditto risk of contamination.

The basic mode of operation of the MWCO filters relies on assisted passage of molecules below a certain size/mass by means of centrifugation. In practice, one adds sample on top of the filter and spins it to induce several thousand times the natural gravity, causing the whole sample to move along the vertical axis of the cartridge and into the membrane. The membrane supposedly retains everything larger than

the specified pore size (the “molecular weight cut-off”-point) and thus produces a filtrate containing only sub-MWCO molecules.

In the study performed by Hölta *et al.* (2012) it was shown that the greatest number of identified endogenous peptides resulted from employing a MWCO filter with a cut-off at 30 kDa, which is substantially greater than the mean mass of the peptides (both expected and subsequently identified). Since our goal was to identify as many peptides as possible, the 30 kDa filters were chosen also for our studies. However; since the size of the pores suggested that a number of small proteins could potentially be included in the filtrate we investigated the filtered sample on an SDS-PAGE gel – but could find no trace of any protein or peptide larger than 6 kDa (see figure 16). The cause to this is difficult to speculate on, but it suggests that simply relying on the specified pore size of the MWCO filter may result in errors and that the method requires re-evaluation if one seeks to study another mass-range.

3.3. Solid phase extraction: principles

Biological samples commonly contain a complex mixture of molecules. The sought after analytes, in this case primarily endogenous peptides, are mixed with other CSF constituents such as salts and cell debris as well as proteins, often referred to as interfering matrix components [238]. Since the presence of the interfering components may negatively impact the MS analysis it is crucial to remove as much as possible without losing analytes of interest. Proteins can be removed by various means, such as MWCO-filtration as described in the previous section. However, after the filtration the sample still contains, beyond the endogenous peptides, small organic molecules (*e.g.*, lipids) as well as non-organic salts and other matrix components, which requires further sample treatment to be removed.

Two approaches for selective exclusion of unwanted molecules was evaluated during the course of this thesis work; solid phase extraction (SPE) and turbulent flow chromatography (TFC). Since TFC is also employed to separate small endogenous peptides from the remaining components of CSF, it is described in a separate section below (see section 2.5.2.). SPE takes advantage of the physiochemical properties of proteins/peptides to selectively retain them, a filter that includes or excludes based on other properties than size/weight [238].

A variety of formats are available for the purpose of SPE, as well as a number of properties for retention [239]. In the current case, disposable cartridges filled with a siliceous packing material coated with hydrocarbon chains with an 18-carbon long backbone (C_{18}) which retain molecules based on hydrophobic interaction were employed. The properties of SPE makes it ideal since that components of CSF span a wide range between the extremes of hydrophilicity/hydrophobicity. Salts are generally highly hydrophilic and cell debris highly hydrophobic, while proteins and peptides commonly occupy a more moderate range on this scale, which can be used to selectively capture them [240].

The employed SPE system uses a crude form of reverse phase chromatographic separation; the peptides and proteins as well as cellular debris are retained on the column because their hydrophobic interaction with the stationary phase is stronger than that to the mobile phase, while salts and other small molecules wash over. In the subsequent step a secondary mobile phase with a suitable concentration of organic component is washed over the cartridge and the proteins/peptides are eluted into the new mobile phase. By employing an eluting mobile phase that contains a sufficiently high concentration of organic component to elute peptides, but not high enough to elute cellular debris, the SPE approach allows for removal of a large section of unwanted matrix components from the sample.

3.4. Losses and contaminants

It must be noted that sample preparation in general also means analyte losses. Since the analytes in question, *i.e.*, peptides, come with such a broad range of physiochemical properties, each preparation step inevitably means losing some of them [186, 241-243]. For instance, transferring the sample from one container to another means losing peptides to the surface of the first container, to the surface of the pipet-tip and then to the surface of the second container.

Secondly, the risk of introducing unwanted components into the sample was an ever-present issue, and especially prominent in the lengthier protocols. Since the MS can detect excessively small quantities of any analyte, there did not need to be large or even noticeable accidents to reduce a sample to a literal cess-pool of unwanted molecules in the context of advanced biochemical analysis. What constitutes a contaminant is highly dependent on the studied analyte. The settings of the MS determine what is detectable and, if peptides is the target, anything that resembles a peptide in terms of physiochemical properties can be a contaminant – including off course peptides [244]. In the case of peptidomics and proteomics the most commonly occurring contaminants are synthetic carbohydrate polymers, particularly, in the view of this researcher, polyethylene glycol (PEG) [244].

Hence, while sample preparation is essential in order to detect peptides, the more extensive the sample preparation was, the higher the risk of introducing various contaminants and the risk of losing peptides. For instance, a balance between the optimal concentration of a chaotrope for breaking protein aggregates and the maximum concentration that causes polymer leaching in subsequent steps was painfully determined.

3.5. Reversed-phase liquid chromatography

Chromatographic separation of the sample constituents is a generally useful practice, especially in bio analysis where sample complexity is usually high. Including an HPLC step prior to MS analysis is meant to limit the amount of analyte species which enters the instrument at any given time. Depending on the analyte-composition of a sample, different modes of separation are employed. For the purpose of temporally spacing out biopolymers (such as peptides), which physiochemical properties are as varied as they are complex, in solutions containing hundreds, thousands or even tens of thousands different analyte species it is important to understand that there is no perfect choice in separation method (at least not at this time). That said; various different approaches have proven highly useful in the chromatographic separation of biopolymers prior to analysis, such as ion exchange chromatography (IEC), possible none more so than the method of employing hydrophobic interaction between the column stationary phase and the analytes, known as reversed-phase liquid chromatography (RPLC) [245, 246].

The RPLC principle of action is to retain peptides based on their higher affinity for the non-polar stationary phase during sample loading with a mobile phase of low (a few %) organic content [247, 248]. By subsequently gradually increasing the hydrophobicity of the mobile phase by increasing the concentration of organic component, peptides again shift their affinity but now from stationary to mobile phase. Since hydrophobicity is an inherent peptide property dependent both on constituent amino acids as well as their relative position *and* various external factors, such as pH and temperature [249-251], RPLC is a highly diverse, dynamic and effective means of peptide separation [252].

In the context it also seems relevant to note the separation technique known as HILIC (hydrophilic interaction liquid chromatography) first described by Alpert (1990) [253], which is in practice, if not in

definition, a form of RPLC where peptide retention is based on hydrophilic interaction and elution is performed by decreasing the hydrophobic component [254].

The ability to control the rate of analytes eluting from the column by means of mobile phase gradient allows RPLC to be one of the most, if not THE most, highly resolving chromatographic separation techniques currently available. Implemented on a high- or ultra-high performing liquid chromatography system, RP separation can allow for a system with unsurpassable peak capacity [255], making it a highly useful, particularly for discovery proteomics.

In the current work RPLC was applied both for on-line purposes (*i.e.*, coupled to an analysis instrument, such as an MS) for straight forward separation of CSF peptides, but also for off-line sample clean-up purposes as well as for pre-fractionation to further deconvolute the sample prior to MS analysis.

3.5.1. Sample pre-fractionation

The study in paper I aimed at deep exploration of the CSF endopeptidome, with the explicit intent of discovering as many endogenous peptides as possible. Various steps were taken to achieve this goal but the most prominent was the introduction of off-line (“not connected to any analytical instrument”) chromatographic sample fractionation of the isolated endogenous peptides following MWCO filtration. The most direct advantage of this procedure is to reduce the complexity of the sample, decreasing the chances of peptide co-elution from the LC-column during sample separation, resulting in improved peak capacity of the LC-MS/MS system and potentially allowing for a higher number of identified peptides [256, 257].

Further and parallel to the reduction of sample complexity, pre-fractionation importantly allowed for the use of a six-fold increase of initial CSF volume, from 250 μ L to 1.5 mL. Since the nano-LC trap and/or separation columns determines how much sample material can be analysed in a single LC-MS/MS run, high sample complexity means that the total analyte mass is “spread out” over a large number of analyte species [258, 259]. By reducing the total number of analyte species in a single fraction and simultaneously increasing the *total amount* of those analytes still present, detection of individual analytes was improved, particularly benefitting detection of low-abundant ditto [229, 256-258].

The obvious disadvantage of fractionation is an increase in the number of analysis runs and thereby a dramatically prolonged total analysis time. To tackle this problem, a strategy for reducing the total number of fractions without causing analyte losses or reintroduce a high degree of complexity was implemented. The solution was to combine fractions of peptides eluting sufficiently far apart, *i.e.*, with sufficiently large difference in eluting hydrophobic mobile gradient component, in a so called “concatenation scheme”. The approach allows analyses of several fractions per LC-MS/MS run, saving time by avoiding for instance sample loading time and column cleaning, and results in a number of “peptide islands” being analysed; peptides from one fraction are eluted and analysed over a period, followed by a pause in peptide elution until the peptides from the next fraction start eluting.

However, the fractionation/concatenation approach can be made more advanced, and the peak-capacity can be further improved, by implementing different LC-separation properties between the off-line and on-line LC steps [260]. In the current case we chose to perform the separation steps over two RPLC columns based on hydrophobic interaction but at two distinct LC-gradient pHs; first off-line at pH 9 followed by on-line at pH 3. Since peptide retention properties can be affected by local pH, this strategy meant that the peptides deposited in one particular fraction at a high pH would slightly alter their retention in the next separation step (on-line at low pH) and elute at comparatively different gradient

point (compared to the off-line separation)[251, 261]. By ensuring, that peptides eluted relatively far apart, but also more or less continuously, high peak-capacity and efficient analysis can be achieved [251, 261-263].

3.5.2. Turbulent flow chromatography

Mass spectrometric instruments for highly sensitive, specific and rapid proteomic analysis is becoming more and more prevalent, and is also available at evermore reasonable prices. The current goal for several manufacturers seems to be to develop “plug-and-play”-type instruments, which requires little in the ways of operation experience to yield results [264-266]. The advantages of employing such MS systems in routine clinical analysis is becoming clearer as more and more protein and peptide biomarkers emerge [5, 12, 267-270]. In paper II we investigated a chromatographic setup for peptidomics meant to bridge the gap between sample pre-treatment and MS-analysis by providing both selective peptide isolation and sample clean-up in a single chromatographic step. The approach is known as TFC and is based around a separation column built to handle comparatively high flow rates whilst otherwise retaining the same functions as a standard reverse phase column based on hydrophobic interaction [271, 272].

In essence, the TFC method employs the physical characteristic that smaller/lighter molecules are more prominently affected by current flow than are larger ditto [273]. Radial motion in tube flow is greater for smaller molecules, resulting in them taking a more indirect (more winding) path through the chromatographic separation column, increasing the chances of the molecule coming in contact with, and being retained on the stationary phase [274, 275].

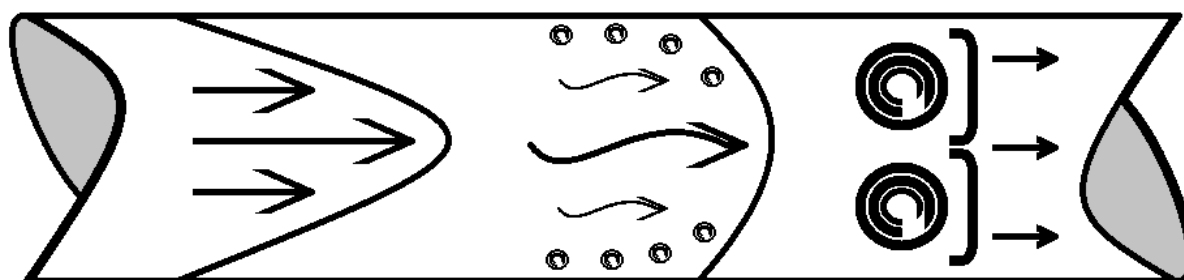


Figure 4: Flow-profile evolution in a closed pipe in response to increasing flow-rate. From left to right: low or laminar flow results in a parabolic flow-profile with the highest relative velocity in the centre of the tube; as the flow-rate increases the comparatively rough inner surface of the tube causes eddies to form and

Briefly, the general protocol for TFC-based isolation of endogenous peptides is carried out as follows. After extraction through lumbar puncture, CSF is briefly pre-treated to denature protein aggregates and subsequently diluted and loaded in the normal-flow liquid chromatograph. The chromatographic instrument (liquid chromatography, LC) employs a high flow-rate (0.5-2.0 mL/min) during sample loading to induce turbulence in the mobile phase and ensure sample liquid bulk turn-over, the flow-rate is then reduced (~0.1 mL/min) to minimise the eluting volume. No particular mobile phase gradient is employed; the sample is loaded in a mobile phase containing a low concentration of organic solvent (*e.g.*, 2-3% acetonitrile) and eluted by a relatively brief pulse (~1-2 min) with increased ditto (50-80% acetonitrile).

In short, TFC is meant to allow for selective retention and isolation of small analytes, *e.g.*, peptides, while larger molecules, *e.g.*, proteins, washes over during sample loading. With some alterations to the

LC-setup described in paper II, the TFC could be connected to a suitable MS for rapid targeted analysis, for instance a triple quadrupole, in a way similar to that described by Chassaing *et al.* in 2000 [276]. Employing TFC in this manner could allow for a highly automated endogenous peptide biomarker assay based on chromatographic sample preparation and MS-analysis.

3.6. Mass spectrometry

In essence, MS means exposing the mass analyser to suitable analytes (ions in gas phase), accelerating the analytes over magnetic or electric fields into a detector and studying the resulting collision to yield a mass-to-charge (m/z) value of the analyte [277]. Albeit the concept is simple the implication is that, with suitable equipment and knowledge, the contents of any sample can be determined, considered useful by some.

The general applicability of MS technology has led to branching and specialisation of instruments for the study of a great variety of analytes in an equally varied set of environments, from the security terminal at the airport to the surface of Mars [278-281]. In the fields of life science, developments in MS technology have focused on dealing with the unique set of problems involved in studying samples containing tens of thousands of large, delicate and intricate molecular species [15]. Further, biological sample materials - whether it being tissue, plasma, CSF, faecal matter *etc* – often constitute a complex array of molecules, known as the sample matrix [282-285]. Most bio analysts focus on a subset of said molecules, for instance lipids, metabolites or proteins. Even so, after isolating the preferred subset of analytes, the samples are still commonly highly intricate and further extensive sample preparation is required before analysis. Finally, the size and physiochemical characteristics of biological molecules often means that they do not readily acquire the features required to be detectable in an MS, demanding specialisation of the instrument for this purpose [17, 22, 282].

In practice, a mass spectrometer consists of three distinct parts; the ion source, the mass analyser and the detector. The ion source constitutes the interface wherein sample analytes acquire a charge, are ionised, and are volatilised (transferred into gas phase) thereby allowing for detection by the mass analyser [286]. There are a number of approaches for the purpose of turning atoms and molecules into ions and the method employed is chosen based on the nature of the sample [287]. Biological molecules tend to be large and easily disrupted and therefore demanding of careful and deliberate handling [24, 288].

Next after the ion source is the mass analyser which determines the area of application of the instrument. Apparent from the name, this is where the ions are “measured”, and data collected and subsequently exported in the form of m/z [289]. As with the mode of ionisation there is several distinctly different mass analysers as well as a number of ways in combining them into so called hybrid instruments [289, 290].

The final stage of the MS is the ion detector which adds a relative intensity, or relative abundance, value to the m/z of each ion species through conversion of the energy in incoming ions into a measurable signal [291, 292]. The detector is essential to the MS instrument and yet the progress in detector development has not kept the same pace as those made in the areas of ion sources and mass analysers. According to Koppenaal and colleagues, writers of one of the few reviews on the subject, this is because of a general lack of appreciation for the complexities involved in choosing an ideal detector as well as a low frequency in specifying what detector is in a particular trial [291]. The fact however, seems to be that because of slow development and lack of differentiation the consumer has not got so many options. One expects that the instrument one has just bought for a few million SEK is equipped with a detector

suitable for the purpose specified, *i.e.*, uniformly detect ions between 100-100 000 m/z , without losing mass response.

3.6.1. Ionisation

Mass spectrometry is based around the more or less predictable behaviour of charged particles in space for assigning specific m/z values to said particles [293]. This means that the instrument is “blind” to non-ionised particles and hence that ionisation of the target analyte is central. Various methods for ionisation have emerged and are employed depending on phase of the sample (gas, liquid or solid) as well as the physiochemical properties of the studied analytes. The MS may have a number of limitations regarding, for instance, analyte mass range but in theory, the mass analyser is able to detect any charged particle that enters it. However, the larger and more complex a molecule is the more difficult it is to ionise and still keep it intact [17].

Early attempts at analysing biological samples were hindered by the difficulties in generating sufficient amounts of ions without destroying the analytes [17, 22, 294]. Ionisation of any molecule/atom occurs as a result of energy transfer to the molecule in question, in the case of large molecules a localised energy surge may cause the molecule to disintegrate [17], and although ms/ms is based around breaking apart (fragmenting) analytes, it has to occur in an orderly manner. For the purpose of studying complex sample materials (*i.e.*, most biological extracts) of large and similarly complex molecules (primarily proteins and peptides), a number of techniques employing so called soft ionisation were developed [16, 17, 21].

“Soft ionisation” in practice means indirect energy transfer and developments of the technique have revolutionised the field of mass spectrometric analysis of bio molecules. Two prominent examples of soft ionisation known as matrix assisted laser-desorption ionisation (MALDI) and electrospray ionisation (ESI), both were developed at the end of the 1980s. In MALDI the sample is transferred to μL wells or “spots” on a target plate and coated with an adsorbing substance (matrix) [295]. The target plate is placed inside the MS under vacuum and the spots are shot at with a UV laser, the effect is complex but results in both the matrix absorbing and transferring incoming energy to the analytes in the sample as well as the sample being “exploded” out of the surface [295, 296].

In comparison, ESI was developed to ionise analytes continuously as they emerge from an LC-capillary, enabling online chromatographic sample separation and mass spectrometric analysis of large biomolecules [16, 17]. As the flow of liquid sample reaches the end of the LC capillary it enters a stainless-steel needle with the nozzle divided by a few cm from the MS-inlet (see figure 5). An electric field in the kilovolt-range is maintained between needle and inlet, resulting in charges being deposited unto the liquid sample droplets as they exit the nozzle [18].

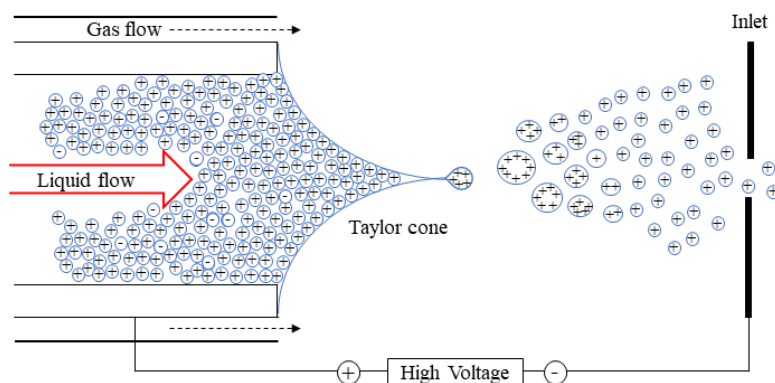


Figure 5: Schematic representation of an electrospray ionisation (ESI) source

As the droplets traverse the interface between nozzle and inlet, the volatile liquid evaporates, concentrating the deposited charges until the charge density is too high and causes the droplet to burst into daughter droplets, a process known as “Coulomb explosion” [17, 18]. This continues until the liquid has evaporated and all that remains (in theory) are single molecules retaining some of the initial charge, *i.e.*, ions in gas phase [17].

3.6.2. Mass analysis

The ways to determine the mass of a particle have been explored extensively resulting in a relatively large variety of different approaches for this purpose. A large number of factors affect what type of sample can be studied by one particular MS (and even more if the instrument is coupled to a chromatographic system). However, the nature of the mass analyser ultimately determines what can be “seen” in a given sample material as well as the quality of the generated data. The fundamental factors of MS analysis, *e.g.*, scan speed, mass range, sensitivity, selectivity, mass accuracy and instrument size affect one another in such a manner that no instrument yet combines the best of all. The complex nature of biological samples has therefore resulted in many life science laboratories employing several MS systems for various types of analysis of a single type of material. Below are described the two mass analysers primarily employed in this thesis work; the quadrupole and the orbitrap.

3.6.2.1. Quadrupole

The basic quadrupole consists of four parallel conductive rods placed, as the name suggests, opposite one another creating a central chamber between in- and outlet (see figure 6). The instrument employs a complicated electric field scheme for influencing ions entering the chamber, which allows for two distinct modes of operation: mass filter or ion trap [297-299].

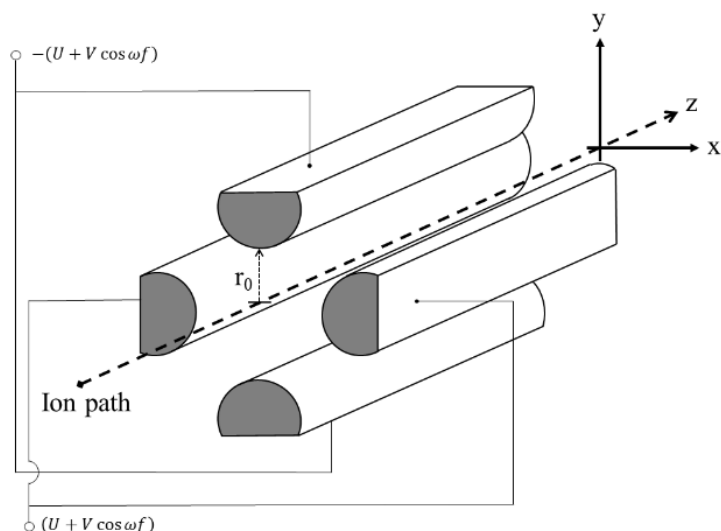


Figure 6: Schematic representation of a quadrupole ion trap

Ions entering the quadrupole are exposed to the effect of a combined electric fields resulting from both a direct current (DC) and an alternating current (AC) induced by a radio frequency (RF) field [298]. The DC field is weak compared to the RF field and its function is primarily as a stabiliser, each rod-pair is consistently either attracting or repelling ions and focusing them in the centre of the quadrupole chamber. The stronger RF field switches polarity periodically, meaning that each rod-pair becomes alternately attractive or repelling causing the ion trajectory to have its own amplitude in both the x and y planes, the RF field thus functions as a destabiliser. Ions that exit the trap reach an external detector and produces a signal. Since individual peptide species are ejected by a unique RF amplitude corresponding to their m/z -ratio, by linearly increasing the amplitude and noting at what point the signal is detected, a specific m/z -ratio and intensity can be given to each analyte, resulting in a mass spectrum [297]. When operating as a mass filter the quadrupole is set to a specific RF amplitude corresponding to a certain m/z -window, the result is that only ions within said m/z -range will have a stable linear trajectory and pass through the filter [300-302].

A common setup is the triple quadrupole (QQQ) MS, *i.e.*, a series of three sequential quadrupoles each with particular tasks which together allows for rapid, robust and sensitive quantitation [190, 303, 304]. The first quadrupole (Q_1) filters for a targeted precursor by implementing a narrow mass window, the second quadrupole (Q_2) is employed as a collision cell to fragment the precursor into product ions and the last quadrupole (Q_3) functions as a second mass filter, this time for one or several product ions (albeit only a single fragment per scan) [299, 301, 304, 305]. The Q_2 , here employed as an ion trap and collision chamber, uses a technique called “momentum-dissipating collisions” (a.k.a. ion cooling) where analyte ions caught in the potential well of the trap are confined to the centre of the quadrupole by low-energy collisions with helium atoms at low pressure [297]. Subsequently the analytes are briefly (~ 10 ms) excited along the z-axis by irradiation at several frequencies simultaneously (multiple-frequency irradiation, or MFI) causing fragmentation of the precursor by multiple collisions with helium atoms in a process known as collision induced dissociation (CID) [306-309]. The product ions are subsequently ejected into the Q_3 where the second mass filter selectively remove all ion species outside the specified m/z -window, allowing only a select set of product ions to reach the detector.

The quadrupole and the triple quadrupole have both kept and found new niches in the analytical sphere, are commonly employed in hybrid instruments, and are much appreciated for their speed, ion storage capacity and, in QQQ-mode, highly sensitive and well adapted for absolute quantitation [310-313]. In

the work performed here the quadrupole was employed in conjunction with an orbitrap mass analyser (see below). Further, pilot trials in developing a rapid biomarker assay on a triple quadrupole were initiated but put on ice due to longer-than-expected method optimisation.

3.6.2.2. Orbitrap

The aforementioned complexity of biological sample materials will not be discussed again, except for the fact that the phenomenon has required and resulted in the development of some extraordinary pieces of bioanalytical technology. The orbitrap mass analyser, introduced for commercial use in the early 20th century by Alexander Makarov and colleagues [314, 315], is an excellent example; combining previously incompatible properties of high resolution, mass accuracy, throughput and broad dynamic range [316, 317].

The mass analyser system is based on the rotational behaviour of ions around an electrode in an electrostatic field (see figure 7) [314]. Ions are injected into the trap at a set energy level and only ions with a suitable tangential velocity will take up a stable orbit around the central electrode. The behaviour is in many ways similar to that of planetary bodies moving around a central star; too low velocity and the orbit decays, making the body fall into the gravitational well of the star, too high velocity and the body bounces out into space (*i.e.*, the inner wall of the trap).

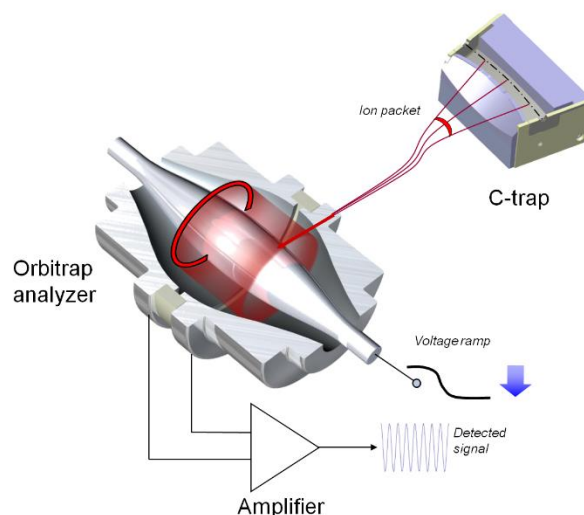


Figure 7: Schematic representation of an orbitrap mass analyser, including the external C-trap. Reprinted with permission from Thermo Scientific, copyright 2018.

The initial tangential velocity and mass determines radial distance of the ion cloud of each analyte species. This characteristic *could* be employed to determine m/z values. However, since the radial distance is affected by initial velocity this analysis approach would be sensitive to variations in the injection energy from the C-trap, resulting in loss of accuracy [317]. Instead the orbitrap employs the phenomenon that the ions orbiting the electrode also moves back and forth, oscillates, along the electrode (z-axis) in a manner dependent only of their m/z -ratio [314]. Because of the harmonic oscillation pattern being independent from rotational and radial velocity, except for the initial requirement of taking up a stable orbit, it can be exploited to determine ion mass with a very high accuracy (>0.1 ppm) [317-319].

Interestingly, the orbitrap is by itself not a particularly well-functioning instrument. For the purpose of continuous analysis of complex samples, in the least it requires an external trap for injecting defined

quantity of analyte ions (the C-trap) as well as some type of collision chamber for product ion fragmentation [265]. Other sought after attributes, not provided by a lonesome orbitrap, is sensitive ion detection and precursor selection [316]. These and other issues were first addressed by simply adding a linear ion trap equipped with a quadrupole ion trap and a CID collision cell upstream of the orbitrap [290]. The LTQ Orbitrap is an example of a hybrid mass spectrometer, intended to combine the best of two worlds; in this case the high resolution and high mass accuracy of the orbitrap and the high sensitivity and ms/ms capability of a linear mass analyser. Since the instruments employed for the vast majority of data collection in this thesis were orbitrap hybrid instruments, this concept will be described in a separate section.

Importantly, the orbitrap mass analyser, combined with the HCD cell (explained in further detail below) as well as a few other features, such as the automatic gain control (AGC), which allows for rapid estimation of the ion population size, has improved the ability to perform relative quantitation through multiplexed isobaric labelling [316].

3.6.3. Ion fragmentation methods

The concept of tandem mass spectrometry involves a primary analysis step in which the incoming ions are detected in the mass analyser and an m/z value is added to each species. For large molecules, such as peptides and proteins, in complex sample matrices, this first step simply does not sufficient determine the exact nature of the molecule – *i.e.*, the amino acid sequence. Hence, a secondary analysis step, an ms/ms, is performed. From the ion stream entering the instrument one selected ion species, a precursor, is filtered to be fragmented, or disassociated, and subsequently having the resulting product ions analysed. The product ion information, the ms/ms data, is eventually employed to infer information on the precursor – this is the essence of tandem mass spectrometry.

There are several modes of fragmentation and the method chosen has a significant impact on the resulting data. In principle the protein/peptide backbone can be broken to primarily generate three distinct pairs of ions: a-x, b-y and c-z (see figure 8). The method of fragmentation determines what type of product ion pair is generated as well as to what extent a precursor is fragmented [320]. Below are the three approaches used for generating product ions for ms/ms-analysis in the works of this thesis described.

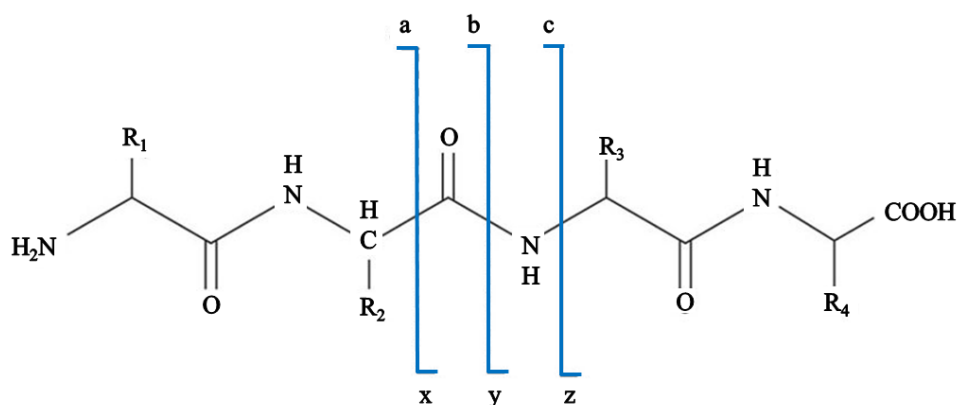


Figure 8: Visualisation of the peptide fragmentation sites along the amide bond backbone, also shown are the three possible ion pairs: a-x, b-y and c-z

3.6.3.1. Collision-induced dissociation

CID and HCD are similar techniques based on, as the names suggest, causing the precursors to break apart through collisions – in both cases with an inert gas, *e.g.*, nitrogen/helium/argon. In both cases (and despite the “higher energy” in HCD) the precursor is typically fragmented through multiple collisions at energy levels between 1-100 eV over a brief period of time [308, 309]. The dissociation of the peptide backbone is not caused by the physical collisions themselves but rather the atomic/molecular vibrations resulting from absorption of momentum energy [308, 309]. The vibrational energy causes an amide bond-break, resulting primarily in complementary b- and y-ions.

The main difference between CID (performed in a linear ion trap) and HCD results from where and under what conditions the collisions are carried out. CID is performed within the confines of an ion trap and the trap thus also must act a collision cell. In contrast, HCD is performed in a dedicated collision cell, usually a higher-level multipole [321]. The dual requirements of the ion trap primarily results in an upper limit to the collision-energy employed in CID, this because a too high energy will cause randomised ejection of ions of all sizes due to instability triggered by transference of momentum [309, 321, 322]. This means that CID has a lower mass cut-off point and cannot detect most immonium ions nor be employed for detecting reporter ions from isobaric labelling reagents such as TMT. HCD, on the other hand, has no lower cut-off mass (except most Orbitrap instruments anyway has a lower limit of 50 Da – but for other reasons) and allows both for improved identification of PTMs as well as reporter ions [321]. Further the higher energy levels allow for a greater degree of fragmentation; that is, HCD generates a larger variety of product ions and thereby higher quality ms/ms data [323].

Even though HCD is generally slower than HCD, the speed of fragmentation makes both methods ideal for LC-MS/MS type experiments where analysis time per peptide is preferably low. Finally, regardless of individual differences both methods have proven instrumental in proteomics and peptidomics. Primarily because the collisional approach is highly efficient in breaking the amide bond-backbone of small to relatively large (<6000 Da) protonated peptides in a quite predictive manner [309, 324-326].

3.6.3.2. Electron transfer dissociation

Electron transfer dissociation (ETD) is a relatively novel method for amide-bond dissociation, discovered to some degree by chance⁴ in the late 90s [327]. Briefly, protein/peptide dissociation is achieved through a three-step process; 1) radical anions are generated by a negative ionisation of (usually) large aromatic hydrocarbons, 2) an electron is transferred from the radical anion to the protonated precursor ion through interaction in the ion trap and, 3) the presence of an odd electron on the otherwise positively charged peptide/protein causes destabilisation and a break occurs along the peptide backbone, generating c- and z-ions as a consequence [328, 329]. Importantly, ETD does not seem to be affected in particular by the size of the precursor but is instead limited by the charge state of the precursor in the generation of product ion pairs [330]. Further, this mode of fragmentation is specific for the peptide backbone and is thus suitable for PTM analysis since ETD does not result in random “knock-outs” of covalent amino acid modification, as is common in CID [321, 330].

⁴ According to the authors they were attempting to combine neutralisation-reionisation with surface-induced dissociation to generate peptide ions for MS (as one does), but found that although the intended approach did not cause any cleavage, the neutralisation reaction with low-energy electrons itself did results in randomised protein dissociation along the amide bond-backbone [327] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 1998, 120, 3265-3266..

The mechanism of ETD is fundamentally different from CID and HCD but still typically/preferably carried out within a slightly modified ion trap or HCD cell [330, 331]. Further, the ions generated differ from those of CID/HCD, the fragmentation is not limited by precursor size and is also well adapted for PTM characterisation. Hence, ETD is introduced more as a complementing method to CID/HCD, particularly useful in top-down experiments [330]. In the Orbitrap Fusion Tribrid MS (previously described) ETD fragmentation is possible to perform in the ion trap while simultaneously running the orbitrap in parallel.

3.6.4. Modes of operation

Depending on the versatility of the instrument available and the conditions of the experiment, the mode of operation can be adjusted to optimise the analysis. A vast array of properties and settings are possible to affect in LC-MS/MS analysis (although many more are outside our ability to influence) and the amount and quality of the data generated is highly dependent on getting them “right” for the particular trial at hand.

When attempting to discover as broad a spectrum of peptides as possible the quality of the data must suffice to allow for successful identification but since quality and time go hand in hand, going beyond the requirements in quality may result in other peptides not getting enough time. At the other end of the operation-mode-span is targeted analysis, where the list of studied analytes (targets) is set from the start and intentionally kept short to allow for recording of as much data as possible on the few – also resulting in enormous losses of potential information, albeit intended.

Below are described the principle modes of operation employed in this thesis, by coincidence they represent the three most common approaches used in proteomics today; discovery, relative quantitation and targeted analysis (*i.e.*, absolute quantitation).

3.6.4.1. Data-dependent acquisition

Data-dependent acquisition (DDA) is primarily employed for discovery (shotgun) proteomics and means that the MS is given a list of priorities determining what precursors will continue on to fragmentation and is in essence aimed at improving data yield [175, 332]. The effect of a number of factors on DDA analysis have been evaluated and are re-evaluated now and again as new software solutions and instruments appear [333]. The main contributors have focused on the adjustable settings with the largest impact on subsequent peptide identification. Reports on the evaluation of resolving power, signal threshold for ms/ms precursor, consecutive ms/ms scans, automatic gain control (AGC) target for both MS and ms/ms (*i.e.*, the estimated amount of ions available), maximum injection time (closely associated with AGC), dynamic exclusion (meant to avoid re-analysing the same precursor over and over) as well as collision energy for fragmentation give a good idea of what to focus on during optimisation [334-337]. However, it is important to note that sample complexity, pre-treatment, the concentration and physicochemical properties of the analytes, chromatographic separation and ultimately variations in each unique instrument (reported for in the case of HCD collision energy) requires a degree of *in house* adjustments too.

If the aim is to achieve a high identification number it is essential that as few peptides as possible go by “unnoticed”, but also that recording time is not wasted on “lost causes”. This means that the MS should perform full scans frequently to pick up new high-quality candidate precursors for ms/ms, which takes time. It is also imperative that each peptide species sent to ms/ms is present in sufficiently high number to be properly detected – meaning the instrument must allow for a certain amount of fill time to collect

the ions, which obviously takes time. Further, each ms/ms-scan takes time, and high resolution takes time; on the other hand, having to re-do the whole experiment because of acquiring low-quality data takes more time than anything.

In conclusion DDA, once optimised for one's particular pleasure, allows for a good deal of instrument automation which is necessary – or essential – for the purpose of successfully performing something so inherently intricate as discovery proteomics.

3.6.4.2. Targeted proteomics

A mass spectrometer is not inherently able to reliably quantify analytes unaided [26, 209]. This shortcoming stems from the need to ionise analytes prior to detection, a quintessential step to the technology [348]. A large number of factors affect how large of a portion of a particular analyte is ionised from a sample. Most important is the physiochemical nature of the analyte itself, but also external factors such as temperature and pressure of the ion source as well as the presence of a large number of different analyte species entering the ion source simultaneously [26, 348]. Hence, the relative intensity/abundance signal says little about the actual concentration of the analyte within a sample, neither does directly comparing intensities of the same peptide between two samples since the sample-to-sample acquisition variation is generally large [178, 338, 339]. The concept of “label-free quantification” should be noted as an important exception since the methods collected under this umbrella term claims to be able to allow for unaided quantification based solely on LC and ms/ms-data [340]. Label-free quantification does however come with its own unique set of issues which requires specialised instrument setups and software to address and several of the approaches are considered controversial, such as the concept of spectral counting [178].

For the reasons mentioned above and despite the noted exceptions, with current technology, attempting concentration determination of individual peptides by means of regular shotgun proteomics can be considered quite difficult [183, 341]. However, the ability to produce synthetic (stable) isotope-labelled (SIL) analogues of peptides has partly addressed this issue by providing absolute quantification through means of comparison to an internal standard. The backside is that the need for a standard makes the approach inapplicable for quantification of large sets of peptides, let alone whole peptidomes [178, 342]. The synthetic peptides will, barring some potential structural issues, have the same physiochemical properties as their native counterparts. Hence, by adding the SIL-peptide to the sample prior to sample pre-treatment, both analytes will co-elute from the LC-separation column, enter the ion source simultaneously and ionise to the same degree [207, 348]. By incorporating one or several amino acids with heavy isotopes of constituent atoms, mainly ^{13}C , ^{15}N and/or ^{18}O , a slight mass-shift between synthetic and native peptide is introduced. By performing targeted analysis (*i.e.*, continuously recording MS and ms/ms data from said analytes over whole or parts of their elution period) on both peptides it is possible to subsequently calculate the concentration of the native peptide by comparison to the SIL peptide [353].

Targeted analysis can be performed in a several ways but here we will mention three common approaches; single reaction monitoring (SRM), multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM). In practice the difference lies in the number of product ions, or transitions, that are monitored, *i.e.*, having their m/z recorded [191, 193, 343]. SRM/MRM analysis is performed in a triple quadrupole, employing the Q_3 to filter either a single transition continuously (SRM) or for several transitions sequentially and repeatedly (MRM) [190, 303] (see figure 9). In PRM analysis the Q_3 is “replaced” by an Orbitrap mass analyser, allowing for monitoring of all product ions continuously [193].

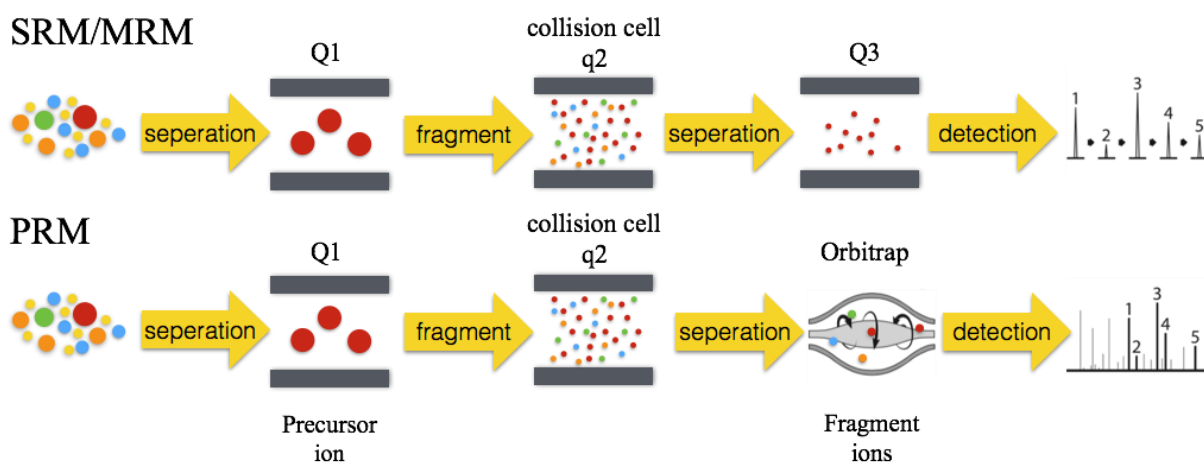


Figure 9: Visualisation of the basic methodology for reaction monitoring employed by triple quadrupole or quadrupole Orbitrap. Reprinted with permission from Creative Proteomics, copyright 2018.

In essence, the difference between the methods lies in speed contra specificity making them suitable for different sample materials. In general terms; SRM focus on speed and transition quality, *i.e.*, generating many data points of a single transition, but is potentially vulnerable to interference (*i.e.*, is not necessarily so specific), PRM on the other hand generates fewer data points because of longer acquisition time, but the monitoring of many transitions with high mass accuracy means PRM is highly specific and has a low risk of interference, lastly MRM is a compromise between the two [190, 193, 303].

Analysis and interpretation of transition data is generally aided by specialised software to generate comparable spectral information through total peak area calculation; that is, the software determines peptide “abundance” by integrating transition intensity over time [189, 344, 345]. By comparing total peak areas of native peptide and SIL standard an abundance-ratio is determined which can subsequently be used to calculate the concentration of the native peptide from the (known) concentration of the standard.

In this thesis I developed a nano-flow RP-HPLC quadrupole-orbitrap (PRM) method for targeted analysis of two endogenous tau peptides, tau 175-190 and its phosphorylated counterpart P-tau 175-190, and their heavy SIL analogues. Data analysis of the acquired transitions was performed using open source software Skyline (v3.7, MacCoss Lab)[189], which was employed to generate native/standard-ratio data from the single most intense transition and using the remaining transitions to verify correct acquisition. Subsequent concentration determination was performed manually in excel (mainly because of a lack of understanding of how to do the same thing directly using Skyline software). The method was employed successfully to generate quantitative data in two clinical cohorts, showing an interesting application for two peptides in AD diagnosis.

3.6.4.3. Relative quantitation

In discovery proteomics it is possible to identify tens of thousands of peptides from a single analysis run and targeted proteomics allows for absolute quantification of a subpopulation of the discovered peptides. There is however a problem in bridging the gap between the two; how to determine which of the discovered peptides should be quantified, which are of potential interest? Finding candidate biomarkers by directly comparing analysis data from discovery-sets between AD and HC will require enormous populations to have even a remote chance of identifying significant alterations (*e.g.*, peptides only present in AD or controls). It is also practically impossible to perform absolute quantification of tens of

thousands of peptides (“exploratory quantification”), and hence one either has to choose targets based on protein association, *i.e.*, “This endogenous peptide is derived from tau. Must be interesting!” or pick them at random or find some other way.

To address the issue of finding quantitative differences between two (or more) states (or individuals) was quite recently introduced a number of peptide labelling techniques that, if employed on the right calibre of MS (high resolving and with high mass accuracy), would allow for relative quantitation of several multiplexed proteomes/peptidomes at a time [346]. The basic MS-analysis method itself is similar to that used for standard shotgun proteomics, so similar in fact that it is shotgun proteomics, but with a sufficiently large twist to merit a separate section. The main difference instead lies in sample pre-treatment and data analysis, with only minor alterations of the DDA-based MS-analysis method [347].

The labelling technique employed during the work in this thesis is known as “tandem mass tag” (TMT) and was first described by Andrew Thompson *et al.* in 2003 [348]. There are other similar labelling reagents available for MS/MS quantitation, such as “isobaric tags for absolute and relative quantitation” (iTRAQ) [349] and the N,N-Dimethyl leucine reagent system (DiLeu) [350]. There are variations in the protocols for the different reagents but the general approach is to separately add the isobaric labelling reagent to up to ten individual samples, covalently attaching the reagent to specific sites on the peptides, quenching the reaction and finally combining the samples (see figure 10).

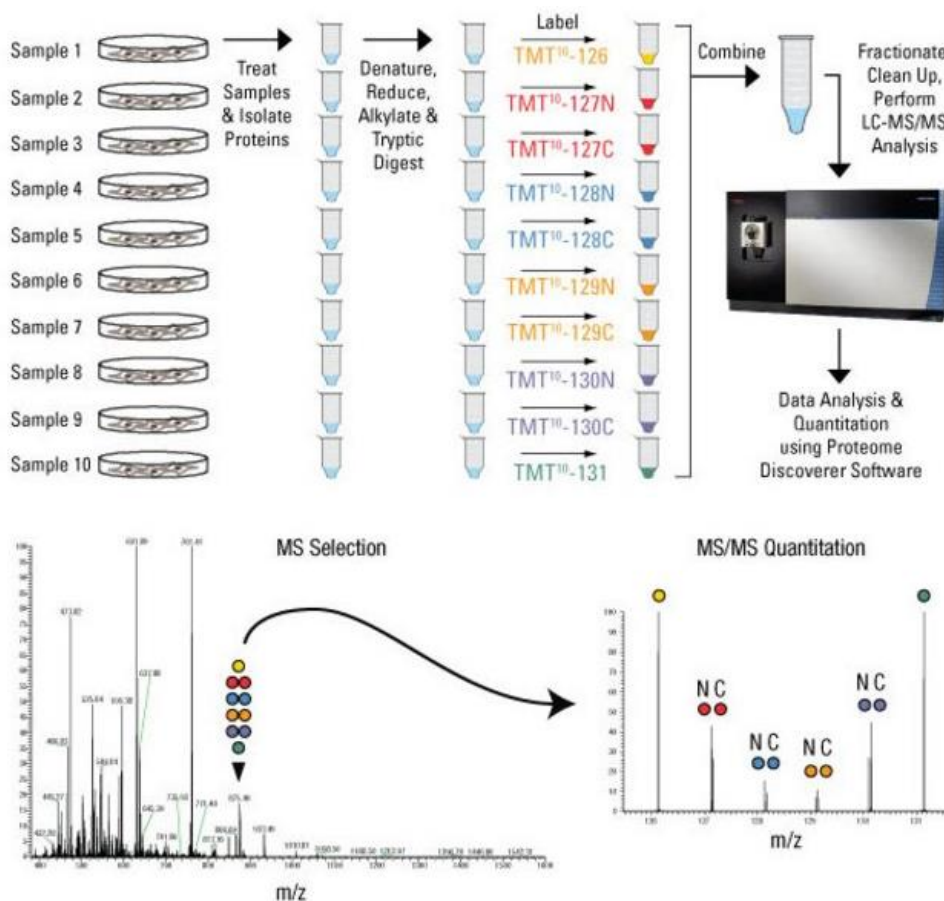


Figure 10: Schematic protocol and workflow employed for relative quantification by means of 10-plexed tandem mass tag (TMT¹⁰) isobaric labelling of peptides from complex biological sample matrices. Reprinted with permission from Thermo Scientific, copyright 2018.

In the case of TMT, the reagent consists of three distinct modules; 1) the amine-reactive group that binds to the peptide N-terminal as well as to free ϵ -amine groups (*i.e.*, on lysine side chain), 2) an isotopic balancing group or mass normaliser and, 3) a reporter group [347, 348]. This elegant solution is based on shifting the load of ^{13}C and ^{15}N heavy isotopes from the balancing group to the reporter group, thereby allowing for ten different isobaric molecules, which will produce reporter ions of escalating mass. The result of individual labelling and subsequent mixing of samples is a single composite peak resulting from the precursor ion, but as the precursor is fragmented in the HCD cell the distinct masses of each reporter ion will result in a ten-step “mass ladder” in the ms/ms spectra with peak intensities corresponding to the relative concentration of the precursor in each original sample. Various alterations to the MS method need to be employed as well as substantial data analysis and interpretation by specialised software is required following LC-MS/MS analysis, but in general terms the price-tag is more complex than the actual application.

The importance of this concept cannot be exaggerated, even though there are of course associated problems – possibly most importantly that one relies too heavily on the generated data [24, 346, 351]. However, the ability to multiplex samples in this manner both improves throughput substantially and may even aid in the detection and identification of low-abundant peptide species [352]. Further and most importantly, the technique has greatly reduced the problems involved in discovery of novel biomarker candidates by providing the means for relatively reliable relative quantification of peptides between individual samples.

3.6.5. Hybrid mass spectrometers

There is a multitude of MS instruments/technologies suitable or specialised for different applications, such as the triple quadrupole MS for rapid, sensitive and robust targeted analysis, the Fourier transform ion cyclotron resonance (FT-ICR) MS for higher resolving and accurate mass-analysis or the previously described Orbitrap which is both fast and high resolving [181, 305, 353, 354]. Because of their distinct capabilities, several mass analysers are often combined in a so-called hybrid mass spectrometer, that combines the particular strengths of the individual analysers and enables specific modes of operation. The vast majority of MSs employ hybrid combinations of mass analysers one of the most common is the Q-‘mass analyser’; that is, a single quadrupole followed by an ion trap and/or mass analyser. The triple quadrupole (QQQ) is one of the earliest examples of this setup, where the last Q work as a second mass analyser (and the middle Q is a trap/collision cell), but other, distinct analysers were soon introduced as well, for instance a time-of-flight (TOF) and subsequently the Orbitrap.

Without getting into too much detail, the quadrupole introduces features such as precursor selectivity and the TOF provides a higher resolution and mass accuracy compared to the “QQQ + detector” setup. This particular setup turned out to be very useful in proteomics since it allowed for continuous high resolving analysis of a complex stream of even more complex ionised analytes, *i.e.*, peptides [22, 355]. Later was introduced the Orbitrap mass analyser which was another “game-changer” since it allows for even higher resolution and mass accuracy [315, 317, 356].

Below follows a brief description of the two instruments employed for all MS analyses performed in the works of this thesis; the Q Exactive MS and the Orbitrap Fusion Tribrid MS, both combining the Orbitrap mass analyser with either a single quadrupole or a complete LTQ. By running mass analysers in tandem, it is possible to combine the features of the two at the expense of instrument size. In both instruments described below the speed and mass filtering capabilities of a quadrupole have been added to the mass accuracy of the Orbitrap. The result are instruments, which may be employed for both exploratory work as well as selective monitoring of specific analytes from complex samples.

3.6.5.1. Q Exactive

The Q Exactive was introduced to be a compact and robust workhorse in proteomic analysis. The Q Exactive is an update or alternative from a previous system, the linear ion trap (LTQ) Orbitrap, which replaces the LTQ with just a Q, a single quadrupole mass filter, and incorporates an HCD cell downstream of the C-trap instead of the CID cell of the LTQ (see figure 11). The setup allows for both high-resolution acquisition with high mass accuracy as well as two modes of targeted analysis, PRM and multiplexed selected ion monitoring (SIM) [357]. Further, the introduction of a separate and dedicated collision cell comes with a few advantages, such as improved fragmentation, ability to monitor low-mass ions and thereby the possibility to perform relative quantification using isobaric labelling.

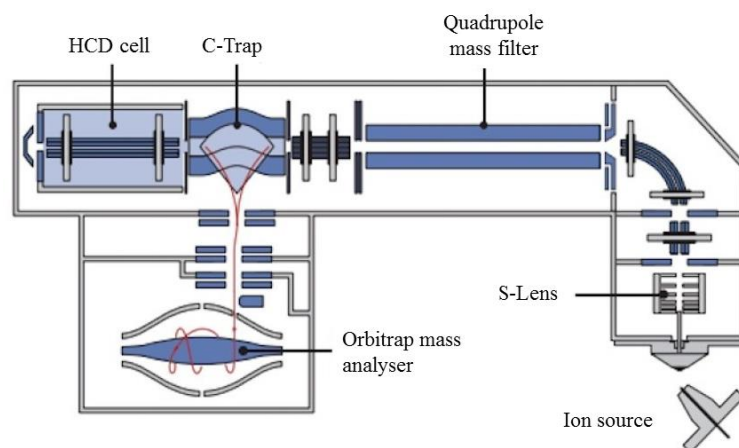


Figure 11: Schematic of the Q Exactive hybrid Orbitrap mass spectrometer. Reprinted with permission from Thermo Scientific, copyright 2018.

Although speed and resolution correlate negatively and a large number of factors impact the actual data quality, the Q Exactive is capable of high frequency MS acquisition (~ 8 Hz) at resolutions around 70 000 FWHM (top resolution ~ 140 000 at 1.5 Hz). The mass accuracy is high also at lower resolution; thus, the resolution in ms/ms mode is chosen depending on the application

Here the Q Exactive was used for high resolution discovery peptidomics of samples prepared by TFC (paper II) as well as for relative quantitative analysis of an isobaric labelled clinical cohort (paper III).

3.6.5.2. Orbitrap Fusion Tribrid

Compared to the Q Exactive, the Orbitrap fusion tribrid (OFT) is substantially larger and comes with an even higher resolving power (maximum 450 000 FWHM), is equipped with a complete LTQ, including a quadrupole mass filter as well as a detector, enabling parallel operation with the Orbitrap mass analyser (see figure 12). Further, the system enables several modes of fragmentation with CID, HCD and ETD/ETHcD (electron-transfer dissociation) thus allowing for an even greater amount of information to be acquired from the analytes, if required. In addition, it took 2 to 3 months “vacation” each summer.

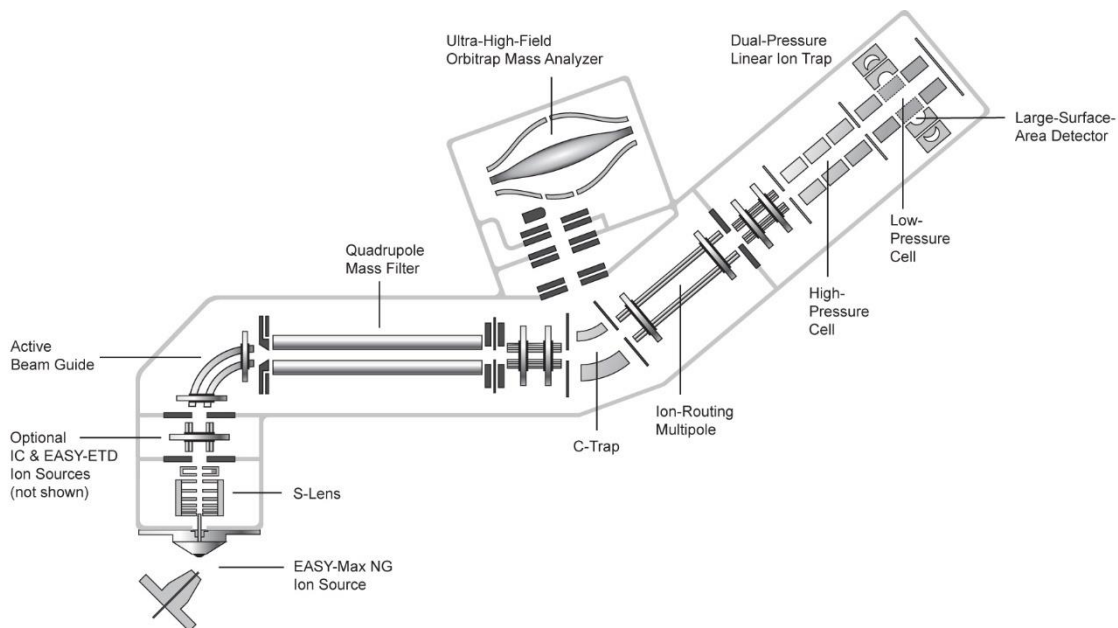


Figure 12: Schematic of the Orbitrap Fusion Tribrid mass spectrometer. Reprinted with permission from Thermo Scientific, copyright 2018.

The OFT was the main analytical tool for most of the work in this thesis and was employed both for discovery and quantitative analysis. In paper I the instrument recorded high quality data in discovery mode, resulting in the identification of more than 18 000 endogenous peptides. In paper III the Fusion was employed both for relative quantitation of an isobaric labelled clinical cohort in data-dependent acquisition mode as well as in targeted top-down analysis; and finally, in paper IV it was used for absolute quantitation by means of a PRM method.

4. Results and discussion

The overall goal of this thesis has been to develop novel biomarkers of NDs, primarily AD. The search for clues to understanding the machinations involved in neurodegeneration has been going on for a long time [358, 359]. Similarly, the sample material chosen was traditional for its purpose, CSF, is for many reasons (see above) considered well suitable for the study of the CNS by proxy. However, the analytes did differ from the mainstream to some extent. Compared to most of the studies carried out for the purpose of investigating the CNS the work performed in the papers described here focused on the endogenous peptide content of CSF.

Of the four papers included in this thesis, the first two primarily involve development of methods for the purpose of acquiring, detecting and identifying endogenous peptides in general. In slight contrast, the last two papers also included targeted analysis of three of the peptides identified as a result of the previous work. In paper I, extensive sample preparation, implementation of state-of-the-art LC-MS/MS equipment and a novel peptide identification strategy was employed to aid in the identification of endogenous peptides in general, and low-abundant ditto in particular. Paper II involved the study and optimisation of a chromatographic technique which would allow for rapid sample preparation intended to be applicable in future targeted analysis of endogenous peptides, for instance in MS-based biomarker assays. The third paper explored a novel approach to endogenous peptide quantitation in discovery proteomics; employing TMT isobaric labelling and spectral analysis based on clustering of the ms/ms data. Paper III also included targeted analysis, first to identify one particularly interesting peptide and subsequently to validate its potential as a biomarker in a clinical cohort. Finally, in paper IV was targeted analysis of two peptides from protein tau, identified in paper I and subsequently in paper II, as the first ever reported endogenous tau peptides performed. The tau peptides were evaluated in two clinical cohorts for their potential use as diagnostic markers for AD as well as for differential diagnosis of AD compared to PD and PSP.

4.1. Paper I: Expanding the CSF peptidome

It was hypothesised that previous studies of the CSF endopeptidome had only scratched the surface and that further discoveries could be made, especially if peptides in the lower concentration ranges could somehow be accessible for detection. As previously noted, one of the major issues in MS-based studies of biological materials is the sheer number of analytes present, and in particular the fact that a few analytes are excessively highly represented [9, 93, 118, 235, 360]. To attempt to overcome this hurdle, extensive sample preparation protocols for pre-treatment, cleaning/de-salting and deconvolution of endogenous CSF peptides for analysis by LC-MS/MS, as well as software-based peptide identification strategies of the resulting data were developed.

A protocol previously developed *in house* by Mikko Hölttä *et al.* (2012) [228] for selective purification of endogenous peptides from CSF, based on MWCO-filtration, was further developed and optimised. Primarily the step for chemical pre-treatment of the sample was modified and off-line RP-HPLC pre-fractionation, based on the work of Tanveer Batth and colleagues (2014) [262], was introduced following peptide purification (MWCO-filtration). On-line RP-HPLC was carried out over a 180 min gradient on an Ultimate 3000 RSLC nano-flow system, allowing for a high degree of peptide separation prior to ms/ms-analysis performed on an Orbitrap Fusion Tribrid mass spectrometer (both from Thermo Scientific). Finally, ms/ms-data was analysed employing a combination of three different peptide identification algorithms. A schematic representation of the workflow has been included in figure 13.

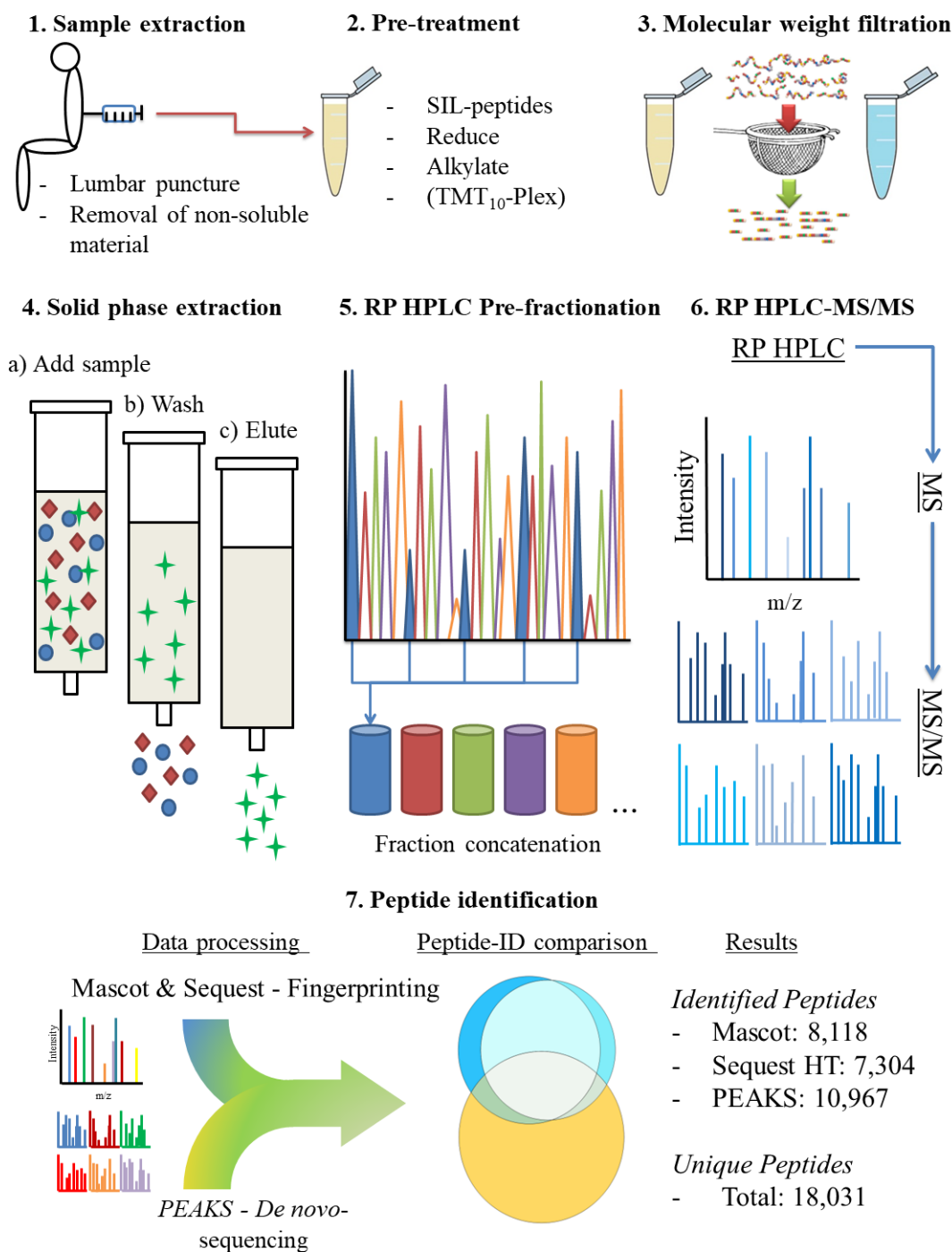


Figure 13: The workflow developed in paper I for attempting identification of as large a section as possible of the CSF endopeptidome. Brief step-by-step explanation; 1) sample extraction by means of lumbar puncture followed by removal of insoluble CSF components by means of sedimentation through centrifugation, 2) chemical sample pre-treatment meant to denature higher protein structures, optional addition of standards or isobaric labelling, 3) removal of proteins through ultrafiltration (MWCO), 4) removal of soluble CSF components and excess reagents as well as small hydrophobic CSF components such as lipids, 5) off-line peptide pre fractionation and concatenation over an alkaline mobile phase gradient, 6) on-line LC-MS/MS analysis of each concatenated fraction, 7) peptide identification through processing of raw ms/ms-data by three peptide identification algorithms, comparison of IDs and evaluation of results.

Since the amount of sample material which can be analysed at-a-time by nano-HPLC MS/MS is limited (this being one of the defining caps for detection of low-abundant peptide species rather than the sensitivity of the instrument) splitting the total peptide content into sub-fractions allowed for an increase

in the total amount of initial sample material used in the protocol [257, 353]. Thus, employing pre-fractionation causes that the relative concentration of each individual peptide species was increased, aiding in the detection of low-abundant peptides [257, 258, 262, 361]. Another effect of fractionating the sample prior to analysis is that the complexity (in practice defined as the number of individual peptide species eluting from the HPLC-column at any given gradient time-point) is reduced, giving the MS a better opportunity to detect each analyte [251, 256, 257].

The last alteration made to the original protocol involved identification of peptides from ms/ms-data. Compared to peptides generated from proteins by means of proteolytic degradation *in vitro* (proteomics), endogenous peptides tend to receive a low identity score when employing proteomic software for ms/ms- data analysis [173, 206, 285, 352]. This issue was partially addressed by engaging a machine learning feature, Percolator™, available in ion-fingerprinting-based software (Mascot, SequestHT), and partially by analysing the data with a tertiary software based on *de-novo* sequencing (PEAKS).

Percolator adapts the scoring algorithm iteratively based on respective common features of the subsets of most and least confident (highest/lowest scoring) peptide sequence match (PSMs) [362, 363]. Since the employed proteomics softwares were originally developed for identifying peptides, generated through proteolytic activity *in vitro*, they may be ill-adapted for endogenous peptides, particularly if said peptides contain PTMs [24]. The percolator feature seemed to alleviate the inherent problems/incompatibilities to some degree when performing *peptidomics* with tools developed for *proteomics* (see figure 14).

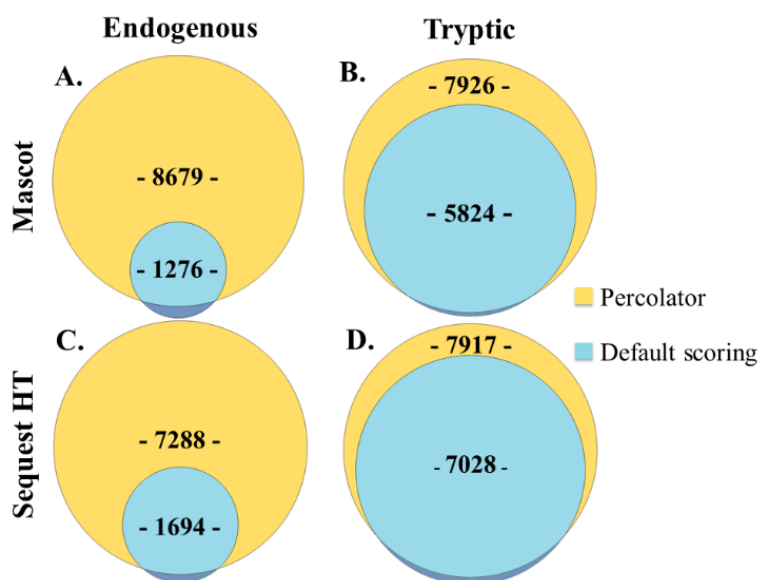


Figure 14: Comparison of number of identified endogenous (A, C) and tryptic (B, D) peptides with Mascot (A, B) and SequestHT (C, D). PSM were scored using the default scoring algorithm of the respective software (blue) as well as the machine learning-based algorithm, Percolator (yellow).

A third algorithm, employed by *de novo*-sequencing software PEAKS, was also engaged to compare the functionality of this strategy to identify endogenous peptides. Very briefly; *de novo*-sequencing algorithms goes through individual ms/ms-spectra and, by measuring distance (in m/z) between detected peaks of fragment ions, sequentially “builds” the sequence one (or a few) amino acids at a time, adding a so called “sequence tag” to each spectra including a confidence value for the particular peptide sequence [364]. Following the initial sequencing the peptides can be compared to a database. However,

the basic concept of *de novo*-sequencing allows for unbiased identification of peptide sequences⁵ and is especially useful for identifying peptides containing PTMs and to pinpoint their most likely position [365, 366].

By further developing the protocol for peptide purification and acquisition and employing pre-fractionation of the sample resulted in a near 10-fold increase in acquired ms/ms spectra during analysis compared to the previous protocol. The 10-fold ms/ms spectra increase, depending on which individual proteomics software was employed (see figure 15), translated to a similar increase in PSMs and a 5- to 8-fold increase in actual peptide identification. Importantly, we were also able to show that it was possible to combine the identification results of the employed softwares without running into multiple testing issues (or we could show that the extent of the multiple testing issues was smaller than the set level of FDR of 1%).

Due to the small total identification overlap, or consensus, between search algorithms of less than 15% we concluded that *not* combining the resulting peptide identification would result in substantial information loss. Hence, we found it reasonable to consider the identification results complementary rather than comparative and combined all unique peptide IDs into a single library of endogenous CSF peptides containing 18.031 entries resulting from three separate analyses of pooled CSF from two different sources (all data available via ProteomeXchange under identifier PXD004863). The approach has been suggested previously by, among others, Shteynberg *et al.* (2013) as an option to optimise data-utilisation in shotgun proteomics [367]. However, we were first in showing the substantial benefit when studying endogenous peptides.

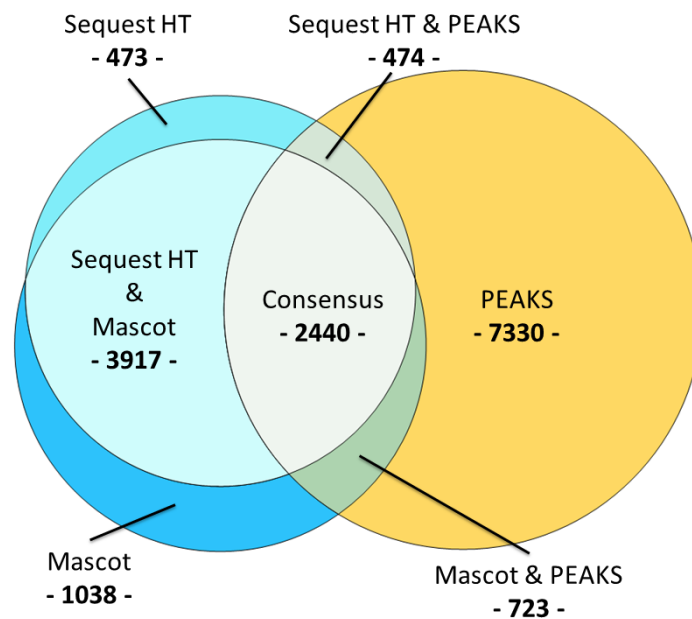


Figure 15: Comparison of proteomics software for identification of endogenous CSF peptides.

Over 1900 proteins were represented by endogenous peptides in the sample set of two CSF pools. Among these were microtubule-associated protein tau (6 peptides), amyloid precursor protein (213 peptides, 58 of which spanning the amyloid beta sequence), NfL and NfH (1 unique peptide each) and

⁵ Ion-fingerprinting algorithms predicate peptide identification on comparison to *in silico*-generated fragment ion libraries which requires pre-existing information on the studied proteome.

a further 60 proteins with known or suspected involvement in neurodegeneration were represented by nearly 3000 peptides [9, 221, 368, 369].

We concluded that the protocol employed here was successful for the purpose of identifying previously undetected sections of the human CSF peptidome. The workflow is labour intensive and substantial sample handling results in increased risks of introducing contaminants. However, this and subsequent studies employing the workflow (see paper III) indicate that investing time and resources in a large scale initial trial may result in a long list of potential candidate markers for further studies. Finally, with the work presented in this paper we were able to show the scale of the human CSF endopeptidome – simply containing such a large amount of information would warrant further study in the opinion of this researcher.

4.2. Paper II: Turbulent Flow Chromatography

In paper II was the potential usefulness of TFC for rapid sample preparation of CSF for subsequent LC-MS/MS analysis of the endogenous peptide content investigated.

As a result of the identification of a large number of endogenous CSF peptides in the previous study (see paper I) it was hypothesised that the CSF peptidome could be a source of future biomarkers for CNS disorders. Biomarker assays based on MS analysis are slowly being considered for introduction into clinical routine, primarily as a compliment to traditional assays based on immunoprecipitation (IP) [13]. There are various reasons for the slow development, but among the major issues is the overall lower throughput of MS-based assays, primarily caused by the extensive sample preparation required [351, 370, 371]. The method investigated in paper II has the potential to address parts of the throughput problem as well as offer a highly automated, specific and sensitive biomarker assays-protocol for future clinical routine analysis based on MS [275, 372, 373].

In paper II we evaluated various factors previously reported to impact peptide acquisition when employing TFC for separation of peptides from proteins and compared the results to the protocol currently in use (*i.e.*, peptide acquisition by means of MWCO-filtration). 500 μ L sample aliquots, corresponding to 250 μ L CSF pre-treated to degrade protein/peptide aggregates, were separated over a Cyclone™, 0.5 x 50 mm, “TurboFlow HTLC”-column (Thermo Scientific) operated offline (*i.e.*, not connected to any analysis instrument) on an Ultimate 3000 normal flow rapid separation liquid chromatography (RSLC) system (Dionex). Following preparation, the resulting endogenous peptides were analysed by nano LC-MS/MS on a Q Exactive (Thermo) system.

Primary factors for testing involved mobile phase pH (3 or 9) and loading flow (0.25, 0.5, 1.0 and 2.0 mL/min) and evaluation of TFC-effect was primarily based on number of peptide IDs resulting from each treatment, as well as ID overlap and variation in a subsequent LC-MS/MS analysis of the sample (see table 1). Further evaluation employed SDS-PAGE gel separation of the protein content of the waste and eluting fractions to visualise the ability to remove the protein content of the CSF at the mentioned flow-rates (see figure 16).

Table 1: Compilation of analysis result from samples prepared by TFC at three different loading flows: 0.5, 1.0 and 2.0 mL/min, and over two mobile phase pH: 3 and 9, and compared to the current standard protocol for separation and acquisition of endogenous peptides; MWCO ultracentrifugation

	pH	Flow-rate [mL/min]	Average peptide IDs	CV [%]	Total # Peptide IDs	ID overlap
TFC separation	3	0.5	936.3	5.4	1526	508 (33.28%)
		1.0	587.0	21.2	968	321 (33.16%)
		2.0	333.0	11.0	462	226 (48.91%)
	9	0.5	1136.0	7.1	1949	555 (28.47%)
		1.0	710.0	10.9	1156	377 (32.61%)
		2.0	465.0	2.8	627	303 (48.32%)
MWCO ultracentrifugation	3	n/a	1354.7	10.1	2260	679 (30.04%)

Initial trials concluded that a flow of 0.25 mL/min resulted in a column backpressure reaching the limits of the LC, causing the LC to shut down or, in several cases, capillaries and fittings to burst. From the PAGE gel it is possible to see that the amount of protein in the fraction passing over the column during loading and elution are roughly similar at 0.25 mL/min. At a flow-rate of 0.5 mL/min, the pressure increase during loading was significantly lower, actually lower than at both 1.0 and 2.0 mL/min as well as 0.25 (data not shown), and no column blocking was noted. This indicates that, for this particular column, the sought-after turbulence of the liquid bulk occurred at some point between 0.25 and 0.5 mL/min, and further confirm that turbulent flow has a dramatic impact on protein—column interaction which has been previously reported [374-376]. From the PAGE gel, it is also possible to see that the protein amount in the wash fraction increases and, reversely, decreases in the eluting fraction with higher flow-rates. These observations support previous findings by Herman and Edge (2012) and others who suggested that TFC can be employed as an adjustable MWCO filter for protein exclusion by altering the flow-rate [377-379].

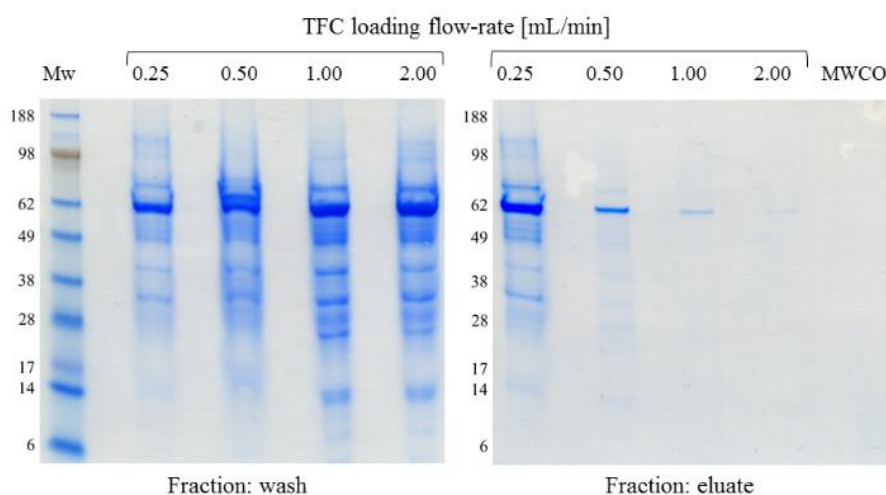


Figure 16: SDS-PAGE gel separation of protein content from fractions collected during TFC sample loading phase (i.e., “Wash” or elution phase, “Elute”, as well as the content of the collection tube following MWCO-filtration); y-axis is presented as an molecular weight “ladder” and the unit is in kDa.

Results further indicate that, once turbulence has been achieved, increasing flow-rate correlates reversely with the number of endogenous peptides identified during subsequent analysis, i.e., the higher the flow-rate the smaller the number of peptide IDs.

Finally, the effect of mobile phase pH is small, but a higher pH seems to correlate with a slightly higher number of identified peptides. However, from figure 17 it is clear that the identification overlap is small and that two different subsets of the CSF peptidome are obtained when applying different mobile phase pHs for TFC separation. The cause is likely a question of peptide properties; that is, some peptides are retained better or worse as a consequence of mobile phase pH [380, 381]. Herman *et al.* (2012) noted that the size exclusion properties of the TFC column at different flow-rates were affected by the pH employed, said results regarded proteins but they may still indicate a potential effect also on peptides [378]. The results indicate that a future assay will have to adapt and optimise the pH of the mobile phase to suit the targeted peptides.

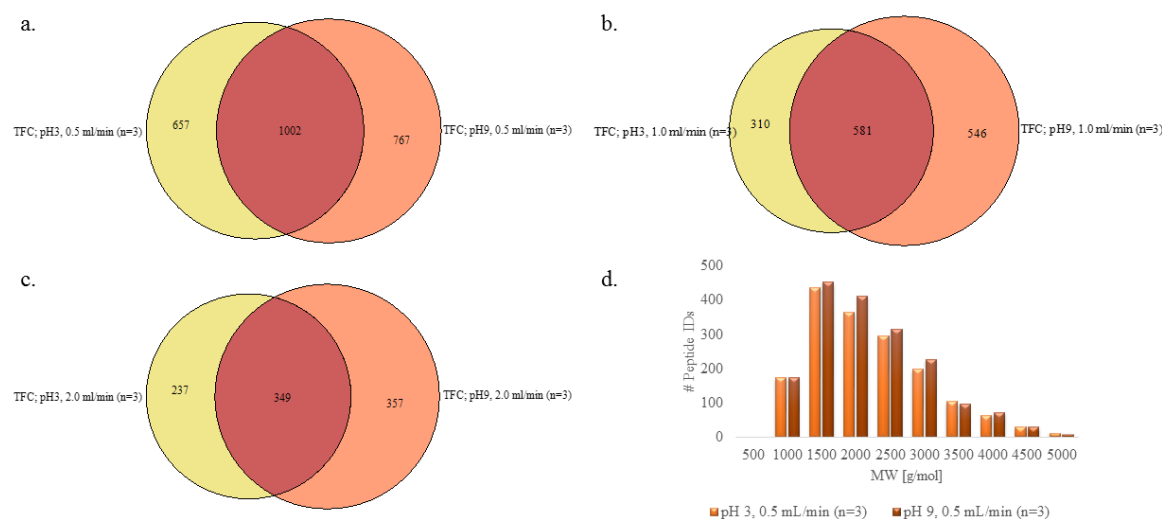


Figure 17: Evaluation of pH-dependence in peptide identification. Venn diagrams showing number of peptide IDs from 3 repeats each from HPLC-MS/MS analyses of samples prepared with a TFC mobile phase gradient at pH 3 or 9 and at flow-rates a) 0.5 mL/min, b) 1.0 mL/min and c) 2.0 mL/min. d) histogram comparing molecular weight distribution of identified peptides from samples prepared by TFC at a flow-rate of 0.5 mL/min at either pH 3 or 9.

The initial results indicate that TFC is applicable for the purpose of selective separation of peptides from the protein content of CSF, comparable to the currently employed method. As previously described, if endogenous peptides can be employed as biomarkers they would be a functional alternative to IP-based assays. As with an IP, an assay based on endogenous peptides does not need to perform a lengthy protein degradation, but measures what is already present in the sample. The current protocol for acquisition of endogenous peptides, based on MWCO-filtration, requires comparatively little time (compared to a standard proteomics protocol) to ready a sample for analysis. Our study shows that a protocol for the same purpose but employing TFC for peptide acquisition would reduce time/sample-requirement further. An LC-setup including a primary TFC step could allow for an (almost) online patient-to-mass spectrometer setup, improving throughput and reducing sample handling [271, 273, 376, 378]. If this increased throughput and automation could be realised the unchallenged specificity of mass spectrometry could properly compete with any IP-based biomarker assay.

4.3. Paper III: Spectral clustering and pleiotrophin

The main goal of the study performed in paper III was, to identify novel endogenous biomarkers for NDs, primarily AD. To this end, the tandem mass tag (TMT) isobaric labelling technique to perform multiplexed (10-plexed) LC-MS/MS analysis of samples of endogenous CSF peptides from two clinical cohorts was implemented. The first, or “discovery”, cohort consisted of a total of 120 individuals clinically diagnosed with AD (n=40) and MCI (n=40) as well as non-demented controls (n=40). The

group MCI was split after a follow-up clinical diagnosis into three sub categories: stable-MCI, MCI-S (n=23), MCI progressed into AD, MCI-AD (n=14) and “other ND” (n=3) – subsequently excluded from the study. The second, “validation”, cohort consisted of a total of 60 individuals clinically diagnosed with AD (n=15), PD (n=15) and PSP (n=15) as well as healthy controls (n=15). Samples were prepared according to the protocol based on MWCO filtration for selective acquisition of endogenous peptides. However, the successful quantitation and discovery of new biomarkers was achieved through the application of a novel proteomic/peptidomics data analysis strategy, meant to address and overcome a severe issue inherent in current identification-driven analysis workflows.

An example of an encounter with this issue can be found in Paper I where we performed a small study on peptide identification success rate. When evaluating ms/ms data in a set of endogenous peptides we found that out of 269 945 ms/ms spectra only 3 094 PSMs were obtained, corresponding to an identification success rate of 1.1%. A similar evaluation of a sample of tryptic peptides gave a success rate of 8.6% (36 886 PSMs from 427 613 ms/ms spectra), which is better, yet the overall conclusion is that a large amount of data does not result in any peptide information.

The problem partly stem from the proteomics algorithms employing protein sequence databases for generating peptide IDs at a given FDR [173]. Peptide IDs are acquired from raw data by comparing acquired MS/MS spectra to *in silico* generated m/z libraries – an approach known as fragment ion fingerprinting [382]. The method is fast but is also necessarily restrictive; peptides from sequences not already present in the database cannot be identified and the number of possible PTMs is strictly limited [383]. Further, peptides not having the correct characteristics, such as a particular amino acid at the correct termini or presenting unsuitable fragmentation spectra, are likewise discarded. The logic is sound: do not waste time and effort on performing quantitation on data that will anyway turn out uninterpretable. However, the quantitative and spectral information can still be employed, but may need further analysis [384].

To overcome this problem, we developed a protocol based on quantitation of clustered ms/ms-spectra prior to actual peptide identification (see figure 18) [384]. The idea being to quantify first, rank the clustered spectra according to how well they separated diagnostic groups and to lastly, attempt to use the ms/ms spectra within the most high-ranking clusters to identify the peptide [384, 385]. Groups of spectra are clustered together based on precursor m/z, charge state and fragment ion patterns [386]. TMT reporter ions from each cluster allows for relative quantitation between individuals in the study groups. This, in turn, permits each individual cluster to be ranked according to ability to separate groups of diagnoses prior to acquiring a peptide ID.

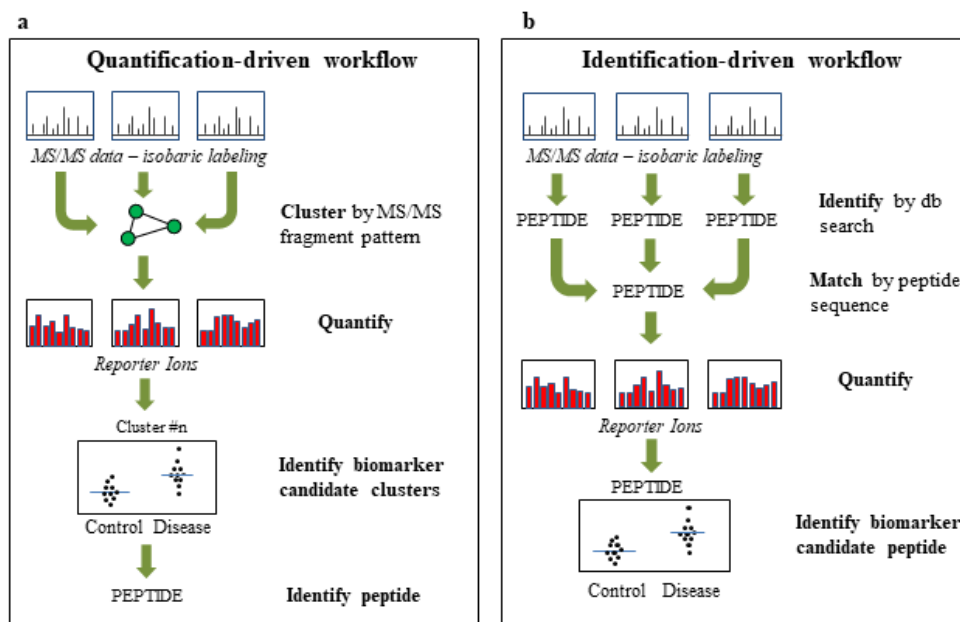


Figure 18: Comparison of the quantitation- and identification-driven workflows. The quantitation-driven workflow (a) initially clusters ms/ms spectra from all TMT sets (and all replicates) based on similarities in precursor m/z, charge state and fragment ion patterns through employment of an algorithm. The TMT reporter ions of individual clusters allow for determination of the relative individual peptide abundance within the cluster and, subsequent ranking of each cluster based on the ability to separate diagnostic groups. Finally, the highest-ranking clusters are identified, either by ion fingerprinting/de novo-sequencing (if possible) or through separate targeted analysis. In the identification-driven workflow (b) ms/ms spectra are initially converted to PSMs and matched by peptide sequence. TMT reporter ions are then used to quantify individual peptide abundances, peptides are ranked according to their ability to separate study groups and finally biomarker candidates are selected.

The result is one or several clusters which separate diagnostic groups sufficiently to merit further study. At this point the ms/ms-data of the fragment ions in the cluster of interest is employed to generate a peptide ID, if possible, which is subsequently used in when attempting targeted analysis for confirmation of the peptide as a potential biomarker. In table 2 is presented the top 20 highest ranking clusters resulting from processing of more than 1.2 million ms/ms spectra from LC-ms/ms analysis (including a total of four replicates) of the discovery cohort. Cluster ranking was based on increasing AUC value (in %) from a receiver operating characteristic (ROC) curve analysis of the ability to discriminate AD from non-demented controls.

Table 2: List of the top 20 clusters in the discovery set, ranked according the calculated AUC of their ability to separate AD and non-demented controls. Column descriptions from left to right: Cluster # (identifier); detected m/z value; detected charge-state; average p-value of cluster similarity to the spectra that are part of a cluster; the number of study subjects in which the cluster was quantified; the relative median difference in abundance between AD and non-demented controls; calculated AUC (from ROC analysis) for discriminating between AD and non-demented controls; successful identification of peptide sequence; the identified peptides' protein of origin.

Cluster #	m/z	Charge	Avg p of cluster similarity	Subjects (n)	Relative diff. AD vs HC	AUC - HC vs AD	Peptide Sequence identified	Protein affiliation
25327	684.044	3	0.003	60	9%**	0.98*	-	
7367	794.115	5	0.073	119	222%**	0.97*	-	
11904	816.467	4	0.001	61	18%*	0.95*	-	
82223	794.885	2	0.000	111	29%**	0.93*	√	Clusterin
61266	1380.71	3	0.000	70	-19%**	0.93*	-	
10464	786.901	4	0.002	59	22%**	0.93*	-	
69078	664.799	2	0.000	94	18%**	0.92*	-	
4243	696.457	5	0.000	59	103%*	0.92*	-	
78065	748.382	2	0.001	120	37%**	0.90*	√	Osteopontin
33238	787.612	3	0.001	69	20%**	0.90*	-	
9084	836.597	5	0.000	86	47%**	0.90*	-	
71623	685.336	2	0.000	61	13%**	0.90*	-	
34935	810.044	3	0.001	58	16%**	0.90*	-	
1889	427.579	3	0.014	120	15%**	0.90*	-	
36033	823.45	3	0.006	120	18%**	0.90*	√	ApoE
87816	848.511	2	0.002	69	18%**	0.89*	-	
6405	690.794	4	0.000	60	13%*	0.89*	-	
32527	777.426	3	0.001	68	13%**	0.88*	-	
13290	840.651	4	0.038	111	18%**	0.88*	√	Secretogranin-1
12243	911.104	5	0.000	70	-28%**	0.88*	-	
53204	1088.54	3	0.000	60	4%*	0.88*	-	

** Indicates statistical significance ($p < .001$)

* Indicates statistical significance ($p < .05$)

The 2nd top most cluster, #7367, was quantifiable in 119/120 subjects and presented a ROC-AUC of 0.97% with a relative median increase in abundance between AD and non-demented controls of 222%. Further, plotting individual intensity ratios of the cluster, grouped according to diagnosis showed an escalating pattern corresponding to disease severity (see figure 19). This suggested a potentially highly useful biomarker candidate for AD. However, initial attempts to identify a peptide sequence failed, as did identification of 15 out of the 19 remaining clusters – thoroughly introducing the backside of this workflow.

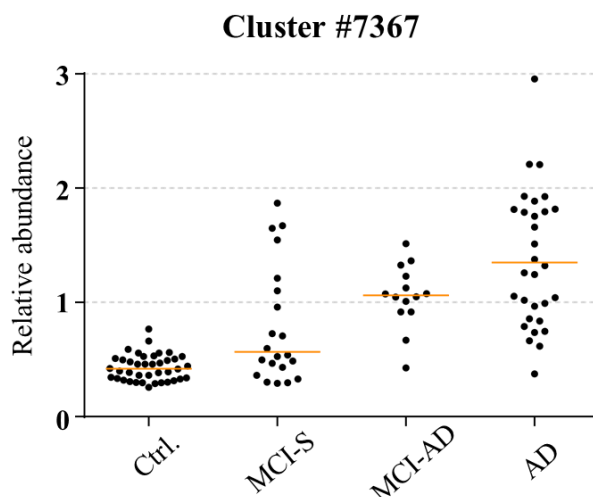


Figure 19: Scatterplot of TMT reporter ion intensity ratio in cluster #7367 from the discovery cohort, grouped according to diagnosis of; non-demented control (Ctrl.), stable MCI (MCI-S), MCI progressed into AD (MCI-AD) and AD.

Making a long, intense and tedious process of analytical detective work (well worth reading about in [387]) into a single sentence; cluster #7367 was eventually identified as peptide sequence 151-AESKKKKKEGKKQEKM-166, of protein pleiotrophin (PTN) with a monoisotopic mass of 3966.54 Da. The PTN protein has been shown to be involved in embryonic and early post-natal neural development as well as in angiogenesis and tissue regeneration [388-391]. However, in later life only certain types of neuronal cells in the hippocampus and cerebral cortex express the protein [392], putting it quite specifically in the middle of the area most prominently affected in AD.

Validation of the endogenous peptide PTN 151-166 as a potential AD biomarker was performed in a secondary CSF material (the validation cohort, N=60), consisting of individuals clinically diagnosed with AD, PD, PSP as well as healthy controls (n=15 of each). Furthermore; a majority of the CSF samples (those with sufficient volume still remaining) were analysed by enzyme-linked immunosorbent assay (ELISA) to determine the concentration (pg/mL) of core AD biomarkers A β ₁₋₄₂, T-tau and P-tau. The CSF samples were prepared in the same manner as the discovery cohort, employing TMT-10 for multiplexed quantitation, but also employing targeted Orbitrap PRM analysis of the +5 charged PTN 151-166 precursor (m/z: 794.115 Da), recording all resulting fragment ions to ensure high quality spectra.

The approach was successful, and results from the targeted LC-MS/MS analysis of PTN 151-166 in the validation cohort are plotted in figure 20. Analysis of the data showed that PTN 151-166 was significantly increased in AD compared to control, however not compared to PD or PSP. Similarly, ROC curve analysis gave an AUC of 0.80 for specificity of PTN 151-166 when separating AD from control, but 0.55 and 0.67 when discriminating between AD/PD or AD/PSP respectively. In comparison to the core AD biomarkers PTN 151-166 proved to be less specific for the purpose of separating AD from controls. However, it should be noted that the peptide turned out to be a rather tricky analyte and further method development could result in improvements of its diagnostic performance. Finally, the fact that PTN 151-166 *does* separate AD from controls with a verified specificity of 80% suggests some involvement in the pathological process and may indicate future applicability in the study of the disorder.

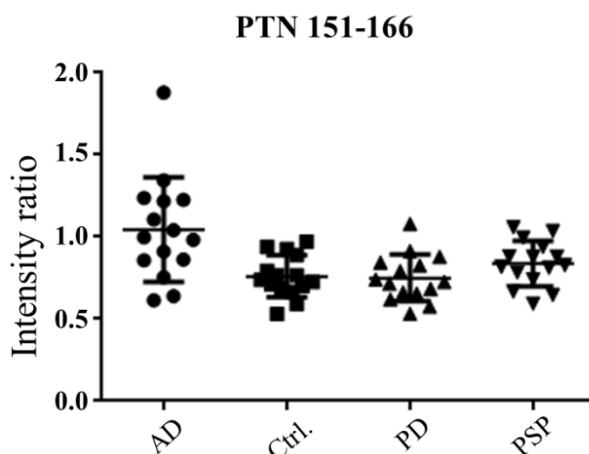


Figure 20: Scatterplot of the relative abundance endogenous peptide PTN 151-166 from the validation cohort, grouped according to diagnosis of; AD, healthy controls, PD and PSP.

4.4. Paper IV: Tau protein-to-peptide ratio

In Paper I was identified the first ever reported endogenous peptides from protein tau; furthermore, several of the identified peptides spanned the analytically interesting site threonine-181 (Thr-181). The CSF concentration of tau (T-tau), measured by ELISA, is a biomarker for general neuronal/axonal trauma [35]. T-tau, along with the CSF concentration of tau phosphorylated at Thr-181 (P-tau) are also used as biomarkers for AD. The identified tau peptides merited further study in order to evaluate whether they could be of further interest as biomarkers for neurodegeneration. To this end a trial with the aim of absolute quantitation of one of the peptides, and its phosphorylated counterpart, in clinical cohorts was conducted.

Endogenous peptide acquisition and purification was performed according to the previously described protocol (paper I) but excluding the pre-fractionation step prior to HPLC-MS/MS analysis. SIL peptides corresponding to their native analogues – the 16 aa long tau 175-190 and its phosphorylated counterpart, P-tau 175-190, were added to the CSF sample as an internal standard of known concentration. An LC-MS/MS-analysis method (PRM) for targeted analysis of the two endogenous tau peptides and their isotopically labelled standard counterparts was developed and optimised on an Orbitrap Tribrid Fusion MS, employing the powerful combination of a quadrupole mass filter and Orbitrap mass analyser. The method was primarily employed for confirming detection and determining concentration of both tau 175-190 and P-tau 175-190 by addition of their SIL peptide counterparts in pooled CSF (see figure 21).

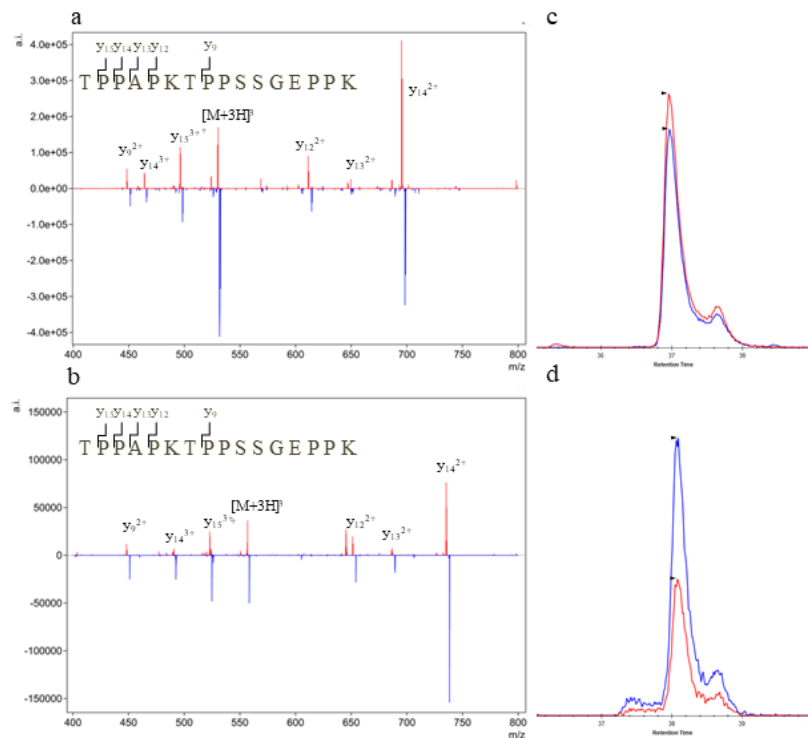


Figure 21: Confirmation of identification and absolute quantitation of tau 175-190 (top row, a & c) and P-tau 175-190 (bottom row, b & d). On the left side are fragment ion spectra of tau 175-190 (a) and P-tau 175-190 (b), both presented in red, visualising similarity in fragmentation with the corresponding SIL peptides (blue) as well as the expected mass shift resulting from the presence of heavy isotopes. On the right are extracted ion chromatograms (XICs) of the precursors of the native (red) and SIL (blue) peptides, meant to visualise similarities in chromatographic behaviour and co-elution.

CSF from a group of healthy control subjects (HC, n=15) as well as individuals diagnosed with AD (n=15), PD (n=15) and PSP (n=15) was prepared and analysed. In parallel was CSF T-tau and P-tau analysed by ELISA on aliquots of the same material, following the standard protocol employed for measuring these core AD biomarkers. Plotting the measured concentrations from the PRM analysis showed no correlation between the concentration of either endogenous peptide and diagnosis (see figure 22 a and b). Further, there was a low or negligible correlation between the measured concentration of T-tau and P-tau protein and their corresponding endogenous peptides (see 22 c and d).

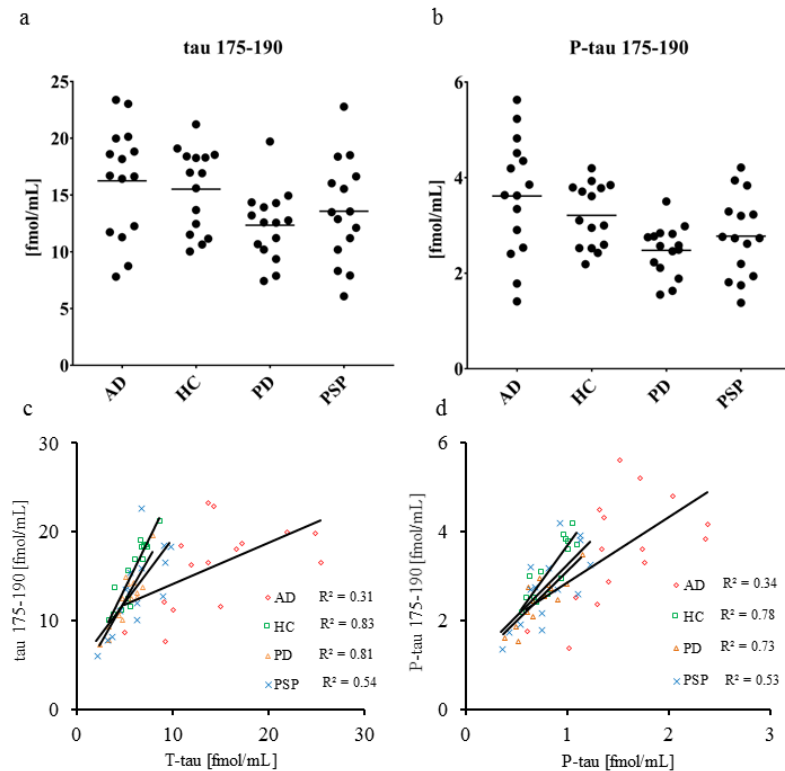


Figure 22: Correlation between total tau protein/total concentration of phosphorylated tau measured by ELISA and corresponding endogenous peptides: tau 175-190 and P-tau 175-190 in fmol/mL

The near complete lack of correlation between proteins and peptides led us to speculate on what origin the endogenous peptides may have. It was possible that tau 175-190 could be the result of normal, non-pathological, processing of tau protein, and P-tau 175-190 conversely the result of ditto processing of phosphorylated tau [115, 122]. If this was the case it might suggest that the endogenous peptides could allow for new insights into the protein. Further, the endogenous peptides could potentially be employed to normalise for the protein expression and improve diagnostic performance. The model for this hypothesis was the reported use of endogenous peptide $A\beta_{1-40}$ to compensate for and reduce variations in $A\beta_{1-42}$ concentration caused by abnormally high/low expression of APP [103, 104].

Figure 23 shows the diagnostic group separation of AD to HC, as well as to PD and PSP, when employing a) T-tau only or b) T-tau/tau 175-190; in c) and d) are corresponding plots for P-tau and P-tau/P-tau 175-190. Although the group separation in this particular cohort is initially good, the protein/peptide normalisation has a notable effect, visualised by ROC curve analyses included in figures 23 e) and f). Further, the effect size for the AD/HC group-separation was significantly increased for T-tau/tau 175-190 (Cohen's $d = 1.98$) compared to T-tau alone ($d = 1.72$); a similar increase in effect size was noted for AD/HC when employing P-tau/P-tau 175-190 ($d = 1.72$) compared to P-tau only ($d = 1.53$) (see supplementary table x).

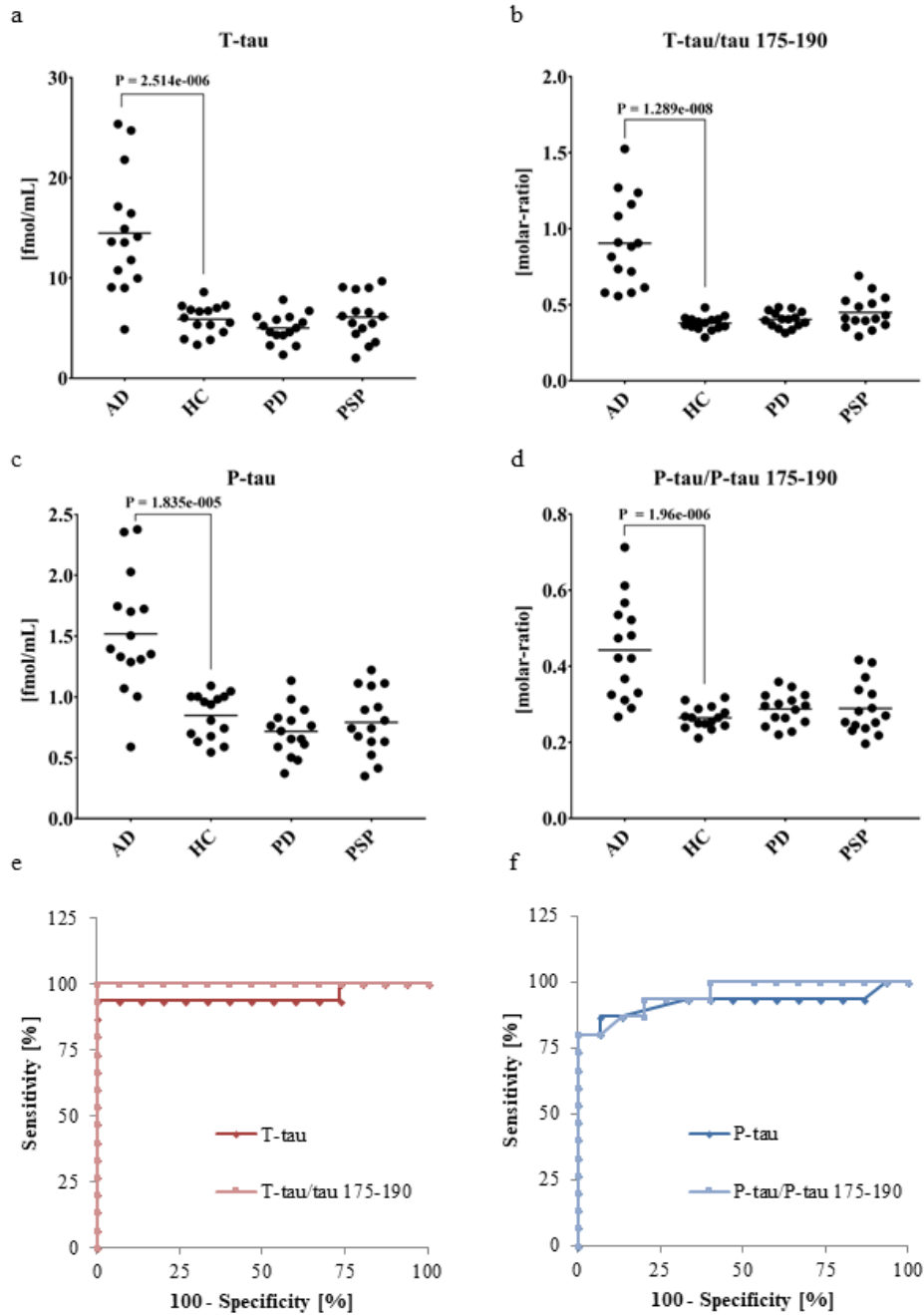


Figure 23: Scatter plots of group separation and ROC analysis of diagnostic accuracy in the discovery cohort. The ability to separate disease groups/healthy controls based on (a) *t*-tau concentrations. (b) *t*-tau/tau 175-190 molar ratio. (c) *P*-tau concentrations and (d) *P*-tau/*P*-tau 175-190 molar ratio. (e, f) ROC curve analysis showing the effect on diagnostic precision when separating AD from healthy controls based on protein concentration alone compared to protein/peptide ratio.

Further study of the set suggested the main effect of employing the protein-to-peptide ratio was in a reduction of group variation, adding to the hypothesis that the endogenous peptides can be employed to compensate for normal expression and processing of tau [393].

A second cohort consisting of 40 clinically diagnosed individuals of AD ($n=16$) and HC ($n=24$) was analysed to study the phenomena further. The results are shown in figure 24 and, albeit (or possibly because of) the groups are less well separated in this cohort, the effect of protein-to-peptide normalisation is clear; group separation is significantly improved.

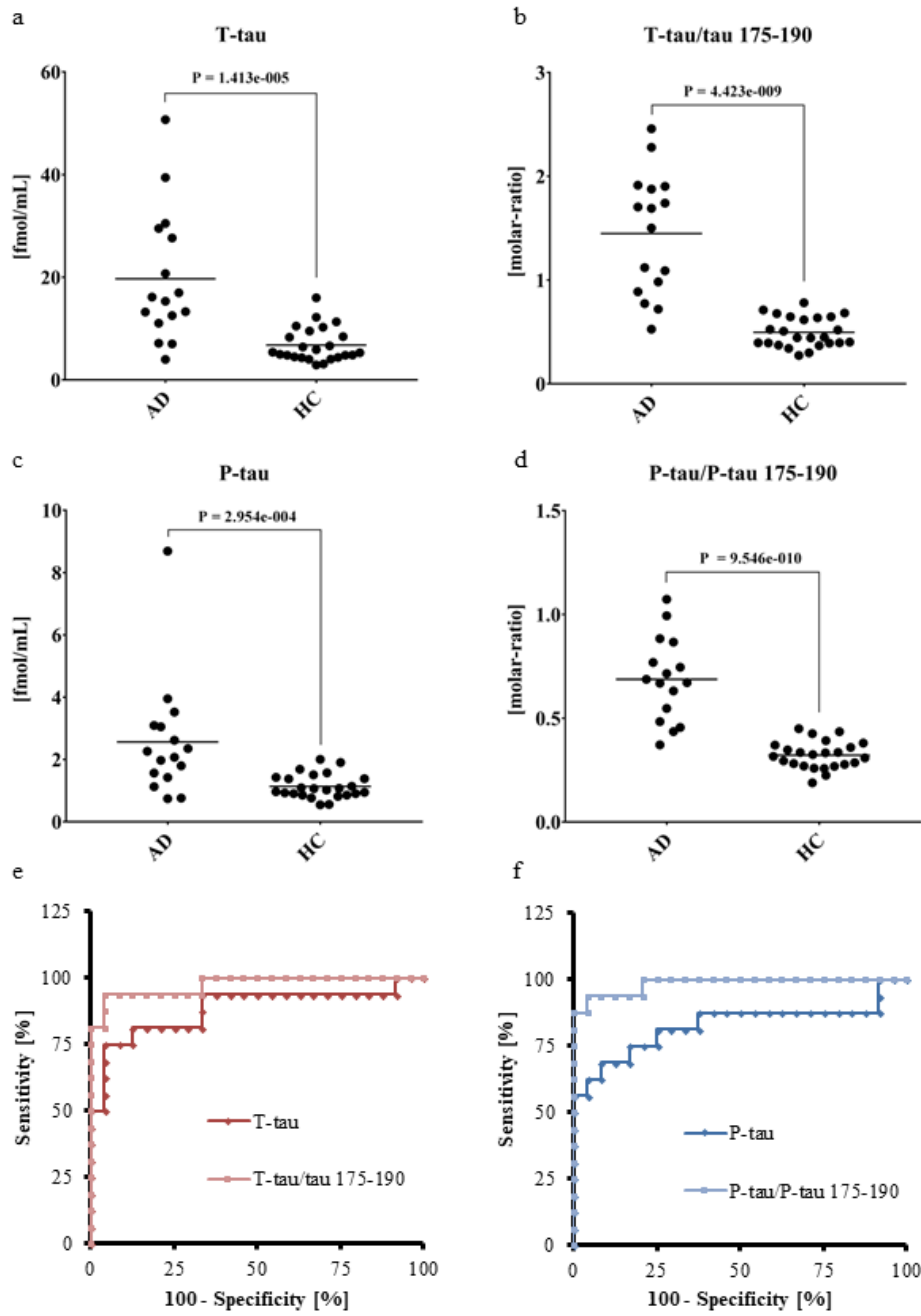


Figure 24: Scatter plots of group separation and ROC analysis of diagnostic accuracy in the validation cohort. The ability to separate AD from healthy controls based on (a) t-tau concentrations. (b) t-tau/tau 175-190 molar ratio. (c) P-tau concentrations and (d) P-tau/P-tau 175-190 molar ratio. (e, f) ROC curve analysis showing the effect on diagnostic precision when separating AD from healthy controls based on protein concentration alone compared to protein/peptide ratio.

5. Conclusions

The major findings presented in this thesis include a substantial expansion of the known CSF endopeptidome, discovery of a number of endogenous peptides from proteins of known or suspected interest in the diagnosis and study of NDs and, introduction of three potentially useful biomarkers for AD. However; most of the present work has been spent developing, adapting and optimising methods for the study of endogenous CSF peptides by means of LC-MS/MS analysis, as well as exploring alternate approaches to peptide identification. Since there is little practical difference between biological sample materials, at least not insurmountable differences, the methods developed for CSF endopeptidomics might with small alterations be applicable for any biological substance, including tissue lysates. Hence, among the main advances presented may be the methods and protocols for the successful study of endogenous peptides.

5.1. The CSF endopeptidome

More than 18.000 endogenous CSF peptides were identified by means of extensive sample preparation and state of the art analysis instrumentation. The most likely reason for the high number of identifications, compared to previous trials, is primarily that the method based on pre-fractionation reduced sample complexity and allowed for an increased total volume of CSF in the sample; both factors combined to improve the ability to study low-abundant analytes.

Further, by employing alternative search strategies, such as combining identifications resulting from three different peptide identification algorithms or performing spectral clustering prior to identification – we avoided losing, or rather, avoided *not using*, a large section of the MS/MS data. Most importantly, we could show that the combination of search algorithms could be performed while keeping peptide IDs resulting from multiple testing within the boundaries set by the employed FDR-level. That is, at the set FDR of 1% one could expect the number of diverging peptide IDs (PSMs) when combining two different search algorithms to be between 0 and 2%, everything above this would be expected to be the result of multiple testing. Since the 2%-boundary was never crossed it was not possible to deduce whether the peptide IDs were the result of multiple testing or a discrepancy between search algorithms. Thus, our approach was within the tolerable margin of error.

The quantitation-driven approach to biomarker discovery, spectral clustering, proved to be equally useful and labour intensive. Although the method needs further development, the advantage of allowing for discovery of quantitative differences in analytes, even with minimal information, is obvious. Since we were able to show that the identity of the peptide in question could eventually be ascertained, we proved that taking this route was useful both as a stand-alone approach and as compliment to traditional discovery proteomics/peptidomics. Possibly one of the greatest advantages however, stem from this strategy allowing for the researcher to be more unbiased since the analyte in question remains unidentified until further analysis. This fact is not to be frowned upon since experience tells me that once the protein origin is established one tends to pick favourites; for instance, everything with the prefix “neuro” makes your heart run a little bit faster.

The ability to explore and quantify the peptides in the lower concentration-ranges of CSF could prove to be the key to discovering the initial signs of neurodegeneration. Although diseases such as AD eventually ends up reducing large parts of the CNS to rubble, the pre-clinical stage is believed to be decades long and the initial pathological process thus most likely leave only faint traces. If AD is ever to be treatable it is necessary to understand these early processes, as well as being able to start treatment

as soon as possible. Since cognitive decline is the main clinical feature of AD, and we so far have no ability to repair a broken brain, a late-stage treatment of AD is simply not appealing. Hence, we believe that the findings and methods presented here may be of use in later studies of NDs and the general study of the CSF/CNS.

A last comment on these discoveries, mainly added because there has been some degree of push-back against the very concept of studying endogenous peptides. We are of course biased in the discussion of the potential usefulness of endogenous CSF peptides, but well-established facts say there is a large number of bioactive endogenous peptides, among them insulin, A β ₁₋₄₂ and various CNS signal substances. Further, a large section of the endogenous peptides discovered in CSF are the results of CNS-processing and are hence on some level indicative of the state of the CNS. Finally, the expanded CSF peptidome presented here has simply not been studied sufficiently to give an indication, positive or negative, of its overall usefulness, only that there are more endogenous peptides than previously shown. Hence; we consider that the belief that the endopeptidome is to be discarded as simply a group of randomly generated waste products to be far from proven. However, if we are in the wrong and this is actually the case, the peptidome would still cause trouble for *proteomics* studies. Particularly those proteomics studies employing quantitation since many of the endogenous peptides present in the sample would look as if they were generated through proteolytic degradation and be quantified as such – thereby resulting in greater or lesser measurement errors.

5.2. The peptides

In paper II and III, we reported on the targeted analysis of the endogenous peptides; tau 175-190, P-tau 175-190 and PTN 151-166. Their usefulness in the diagnosis and study of AD (and to some extent in other NDs) was evaluated in clinical cohorts and compared to and with the current core biomarkers. In the case of the two tau peptides we noted that their implementation as normalisers of their respective protein counterparts could allow for improved group separation and diagnostic accuracy. The PTN peptide by itself was able to separate AD from controls with an 80% accuracy which, albeit less accurate than the core biomarkers, shows potential.

In paper I, we discovered a large number of endogenous peptides stemming from apolipoprotein E, a protein which three most common isoforms have a noted protective, neutral or negative impact on AD development, respectively. The gene coding for the protein constitute one of the few known genetic factors impacting AD and is thus of some interest in the study of the disease. The alteration of two separate amino acids in the protein sequence determine which of the three isoforms is primarily expressed, and thus which role the protein has in the disease. Endogenous peptides spanning both sites were discovered during the exploration of the CSF endopeptidome, and although this study is still on the “to-do” list the implication is clear; genotyping by means of endogenous peptides could be possible.

These examples suggest that, even though the implementation is not always as straight forward, the endogenous peptides can be employed in various ways to ascertain information about the state of the CNS. Since the consensus is that all NDs, and AD in particular, are difficult study subjects with a myriad of complicating factors tied to them it might be necessary to introduce more complex diagnostic approaches. Combining peptidomics and proteomics and taking advantage of the fact that the two yield information on different parts of the whole might do the trick.

6. Future perspectives

From where and from what processes do all these endogenous peptides originate? Suggested sources include the proteasome as a major contributor, but it is possible that various pathological processes can be attributed as well. Further, as has previously been discussed and re-iterated, there are plenty of bioactive endogenous peptides which are the result of normal and highly specific processes. The latter two were at least officially the focus of the initial studies performed in this thesis. The study of the proteasomal involvement in peptide origin could start with observations of media from neuronal cell cultures of induced pluripotent stem (IPS) cell lines, wherein the proteasomal system can be artificially regulated. However; since the number of endogenous peptides is so large it will most likely prove difficult to determine the exact impact of proteasomal regulation. It is worth a shot, but it is quite possible that the origin of endogenous peptides might have to be determined on an individual peptide basis.

Further expansion of the CSF endopeptidome. Although a large number of peptides were identified employing the reported methods, it is reasonable to believe there were groups of peptides that for various reasons remained undetected/unidentified. As was previously discussed, losses occur at every preparation step. The losses are either general, *e.g.*, peptides adhering to the surface of an Eppendorf tube, or specific, as for instance during chromatographic peptide separation when the most hydrophobic/hydrophilic peptides either pass right over the column or are too strongly retained to elute with the gradient. And again, and lastly; not all detected peptides are identified.

In peptidomics, since the peptides are not meant to directly infer information on a protein, every unique identification is potentially valuable in itself. Therefore, are all peptides of potential interest and the CSF peptidome should be further scrutinized. We suggest performing further studies employing several fractionation steps including strong anion/cation exchange (SAX/SCX), isoelectric focusing. Further, the peptides studied here were relatively small and it is possible that bioactive endogenous peptides are generally larger. For this purpose, steps to acquire and detect larger analytes should be taken. Suggestions include trying out MWCO-filters with larger pores (very simple alteration but might have an impact), employ ETD/EThcD as MS-fragmentation method, perform MS/MS on a high resolving TOF-instrument etc.

Implementation of TFC in sample preparation for clinical routine, or at least develop the method so it is ready when there are some endogenous peptide biomarkers available. There are a few hurdles to overcome, such as reducing the concentration of organic component between eluting the peptides from the TFC column and loading them on the separation column. Also, as hinted before, a few biomarkers for the proposed assay would not come amiss.

Lastly, a great number of endogenous peptides were discovered during our trials, and there is both qualitative and quantitative data available on a large section of these peptides. Further, in Paper III we presented a table of 20 clusters with promising quantitative differences between AD and controls. Only five of these peptides were eventually properly identified and only PTN 151-166 was further studied. There might be a slight tendency in omics to get a bit carried away with trying out new ways of generating larger and larger datasets, but the lowest hanging fruit might be to perform a thorough study of the candidates already identified.

Acknowledgements

First and foremost, thank you **Johan Gobom**, for being a great teacher, mentor, friend and supervisor; and for teaching me about filtration, separation, ionisation, quantitation, analysis, detection and identification, as well as about all the “joys” of ROC curve analysis and how to kill darlings. Also, and especially for all the strange, obscene and unproductive conversations that made many of my days. It was a pleasure and a privilege working for you. Special thanks also to my co-supervisor, **Erik Portelius** for sharing both insights and good advice, for always being helpful, for the enormous tub of honey and for running the proteomics section, not only *as* a boss, but *like* a boss.

To our tandem professors; **Henrik Zetterberg** for unrivalled gusto, insight, creativity and a most inspiring love of all things science and **Kaj Blennow** for allowing me to take advantage of your tremendous knowledge into the aggregated nest which is neuroscience, for guidance and your no-nonsense approach to study design – thank you for giving me this opportunity.

Thank you, **Ann Brinkmalm, Gunnar Brinkmalm, Josef Pannee, Kina Höglund, Jessica Holmén-Larsson** and **Jörg Hanrieder** for discussion of all types; scientific, political, historic and weird, for always lending a hand or an ear, for messing up my machines (so I could leave the office for a while), for great ideas and suggestions, for good collaborations and for baby-diagnostics by proxy.

To my PhD-comrades from the start; **Hlin Kvartsberg** for keeping everyone else in check and for teaching me extremely invasive brain surgery, **Tobias Skillbäck** for good collaborations, great talks and for keeping Johan occupied, and **Simon Sjödin** for all the help you have lent me over the years, both in research and bureaucracy – a thousand thanks all three of you for being fantastic company and for listening to my rants... and sorry for writing so loud.

For nearly succeeding to collaborate with me, thank you **Wojciech Michno, Rakesh Banote** and **Faisal Nazir**; it would probably have led to at least one Nobel Prize.

Thank you, **Silke Kern** for letting me tag along for CSF extraction and providing me with samples, **Katarina Tomazin** for taking care of the lab and for keeping my pipets in fantastic condition, **Gösta Karlsson** for keeping everything running and for not beating me when I ruined expensive stuff for the n^{th} time, and **Ulrika Brynnel** for providing some much-needed high quality chocolate and a lot of great input to the final version of this book.

Thank you all members of the proteomics section **Annika Öhrfelt, Celia Hök Fröhlander, Claudia Cicognola, Eleni Gkanatsiou, Fani Pujol, Jessica Wahlgren, Juan Lantero Rodriguez, Karolina Minta, Malgorzata Rozga, Maya Arvidsson Rådestig** and **Patrick Wehrli**.

I would also like to express my gratitude to **Alexandra Abramsson, Alexandra Olsson Ek, Anna-Karin Björklund, Ann-Charlotte Hansson, Anni Westerlund, Annika Lekman, Astrid Marianne von Euler Chelpin, Bob Olsson, Bruno Becker, Carina Molin, Charlotta Otter, Charlotte Frick, Chatarina Andersson, Christoffer Rosén, Dzemila Secic, Ekaterina Lind, Elena Camporesi, Emma Sjons, Erik Hardselius, Frida Gustavsson, Hanane Belhaj, Ibrahim Kaya, Ingrid Maria Englund, Irina Nilsson, Jasmine Chebli, Jennie Larsson, Karin Palm, Kerstin Andersson, Khalegh Adelzadeh, Kristina Sernestrand, Lena Olvén Andersson, Lobna Almasalmeh, Lotta Agholme, Marcus Nordén, Maria Holmström, Maria Lindbjær Andersson, Mariann Wall, Marie Kalm, Mathias Sauer, Monica Malmberg, Payel Bhattacharjee, Petra Bergström, Rita Persson, Rolf Ekman, Ronald Lautner, Rose-Marie Fishman Grell, Sara Skoglar, Sheila Engdahl, Sofia Rasch,**

Staffan Persson, Tove Hellkvist, Tugce Munise Satir, Ulf Andreasson, Ulla-Stina Danielsson, Ulrika Sjöbom, Ulrika Wallström, Victor Liman, Åsa Källén and Åsa Sandelius for running the labs, seeing to the paperwork, sciencing with me and for making the department a great place to work! Special thanks to **Bozena Jakubowicz-Zayer** for showing me around and explaining the importance of smiling at people in the corridor.

Thanks to all collaborators and co-authors not already mentioned; **Elin Pernevik, Oscar Hansson, Ross Paterson, Jonathan Schott, Nadia Magdalinou, Niklas Mattsson, Ekaterina Mirgorodskaya, Wiesje van der Flier, Philip Scheltens, Floor Duits and Charlotte Teunissen.**

Further, thanks to the Swedish Research Council, Wallströms and Sjöbloms Stiftelse, Gun and Bertil Stohnes Stiftelse, Magnus Bergwalls Stiftelse, Åhléns-stiftelsen, Alzheimerfonden, Demensförbundet, Stiftelsen för Gamla Tjänarinnor, Knut and Alice Wallenberg Foundation, Frimurarestiftelsen, and FoU-Västra Götalandsregionen for funding my research.

Thank you **Rahil Dahlén**, you have been (and are) a great friend, guru and teacher as well as an average student. You have helped me tremendously and have been the most positively annoying person I have ever met.

Thank you, **Elin Bergman**, for psychological evaluation and for the cover illustration

Tack **Adrian** 

To my wonderful parents, **Elisabeth** and **Tommy Hansson** (now better known as **Mammo** and **Vavva**) for your unlimited support, for sharing your vast scientific knowledge and experience, for listening to my rants and especially for spoiling my baby when I was too busy.

Alexandra Bergman, the second greatest love of my life, you have helped me in more ways than you can imagine. You have carried me and our little family while still relentlessly pursuing your own scientific goals; you impress me and inspire me and provide comfort and insights without which I could not have finished this book.

To **Ingrid Bergman**, the greatest love of my life. I hope you will read this book in a decade or two (at which point I have probably forgotten about it) so that you may learn a little bit about mass spectrometry, but also so that you may know that the eight months I spent at home with you – while also writing this book – was the best time of my life (despite the book writing).

References

- [1] Reitz, C., Brayne, C., Mayeux, R., Epidemiology of Alzheimer disease. *Nature Reviews Neurology* 2011, 7, 137.
- [2] Scheltens, P., Blennow, K., Breteler, M. M., de Strooper, B., *et al.*, Alzheimer's disease. *Lancet* 2016, 388, 505-517.
- [3] Portelius, E., Brinkmalm, G., Pannee, J., Zetterberg, H., *et al.*, Proteomic studies of cerebrospinal fluid biomarkers of Alzheimer's disease: an update. *Expert review of proteomics* 2017, 14, 1007-1020.
- [4] Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H., *et al.*, Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of general psychiatry* 2003, 60, 709-717.
- [5] Blennow, K., Zetterberg, H., Understanding biomarkers of neurodegeneration: Ultrasensitive detection techniques pave the way for mechanistic understanding. *Nature Medicine* 2015, 21, 217-219.
- [6] Filiou, M. D., Turck, C. W., Martins-de-Souza, D., Quantitative proteomics for investigating psychiatric disorders. *Proteomics. Clinical applications* 2011, 5, 38-49.
- [7] Guest, P. C., Martins-de-Souza, D., Schwarz, E., Rahmoune, H., *et al.*, Proteomic profiling in schizophrenia: enabling stratification for more effective treatment. *Genome medicine* 2013, 5, 25.
- [8] Parker, C. E., Pearson, T. W., Anderson, N. L., Borchers, C. H., Mass-spectrometry-based clinical proteomics--a review and prospective. *Analyst* 2010, 135, 1830-1838.
- [9] Huhmer, A. F., Biringer, R. G., Amato, H., Fonteh, A. N., Harrington, M. G., Protein analysis in human cerebrospinal fluid: Physiological aspects, current progress and future challenges. *Disease markers* 2006, 22, 3-26.
- [10] Schutzer, S. E., Liu, T., Natelson, B. H., Angel, T. E., *et al.*, Establishing the proteome of normal human cerebrospinal fluid. *PLoS One* 2010, 5, e10980.
- [11] Blennow, K., Hampel, H., Weiner, M., Zetterberg, H., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature reviews. Neurology* 2010, 6, 131-144.
- [12] Olsson, B., Lautner, R., Andreasson, U., Ohrfelt, A., *et al.*, CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol* 2016, 15, 673-684.
- [13] Lewczuk, P., Riederer, P., O'Bryant, S. E., Verbeek, M. M., *et al.*, Cerebrospinal fluid and blood biomarkers for neurodegenerative dementias: An update of the Consensus of the Task Force on Biological Markers in Psychiatry of the World Federation of Societies of Biological Psychiatry. *The world journal of biological psychiatry* 2018, 19, 244-328.
- [14] Magdalinou, N., Lees, A. J., Zetterberg, H., Cerebrospinal fluid biomarkers in parkinsonian conditions: an update and future directions. *Journal of neurology, neurosurgery, and psychiatry* 2014, 85, 1065-1075.
- [15] Domon, B., Aebersold, R., Mass spectrometry and protein analysis. *Science* 2006, 312, 212-217.
- [16] Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., Udseth, H. R., New developments in biochemical mass spectrometry: electrospray ionization. *Analytical Chemistry* 1990, 62, 882-899.

- [17] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.
- [18] Whitehouse, C. M., Dreyer, R. N., Yamashita, M., Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Analytical Chemistry* 1985, *57*, 675-679.
- [19] Cornett, D. S., Reyzer, M. L., Chaurand, P., Caprioli, R. M., MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nature methods* 2007, *4*, 828-833.
- [20] Caprioli, R. M., Farmer, T. B., Gile, J., Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Analytical chemistry* 1997, *69*, 4751-4760.
- [21] Verhaert, P., Uttenweiler-Joseph, S., de Vries, M., Loboda, A., *et al.*, Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: an elegant tool for peptidomics. *Proteomics* 2001, *1*, 118-131.
- [22] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198-207.
- [23] Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., *et al.*, Peptidomics. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 2004, *803*, 3-16.
- [24] Dallas, D. C., Guerrero, A., Parker, E. A., Robinson, R. C., *et al.*, Current peptidomics: applications, purification, identification, quantification, and functional analysis. *Proteomics* 2015, *15*, 1026-1038.
- [25] Hökfelt, T., Millhorn, D., Seroogy, K., Tsuruo, Y., *et al.*, *Regulatory peptides*, Springer 1989, pp. 154-179.
- [26] Association, A. P., *Diagnostic and statistical manual of mental disorders (DSM-5®)*, American Psychiatric Pub 2013.
- [27] Wimo, A., Guerchet, M., Ali, G. C., Wu, Y. T., *et al.*, The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2017, *13*, 1-7.
- [28] Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., *et al.*, Defeating Alzheimer's disease and other dementias: a priority for European science and society. *The Lancet Neurology* 2016, *15*, 455-532.
- [29] Prince, M. J., Wu, F., Guo, Y., Robledo, L. M. G., *et al.*, The burden of disease in older people and implications for health policy and practice. *The Lancet* 2015, *385*, 549-562.
- [30] Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., *et al.*, Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of neurology* 1991, *30*, 572-580.
- [31] Petzold, A., Neurofilament phosphoforms: surrogate markers for axonal injury, degeneration and loss. *Journal of the neurological sciences* 2005, *233*, 183-198.
- [32] Kester, M. I., Teunissen, C. E., Crimmins, D. L., Herries, E. M., *et al.*, Neurogranin as a cerebrospinal fluid biomarker for synaptic loss in symptomatic Alzheimer disease. *JAMA neurology* 2015, *72*, 1275-1280.
- [33] Williams-Gray, C. H., Worth, P. F., Parkinson's disease. *Medicine* 2016, *44*, 542-546.
- [34] Bridi, J. C., Hirth, F., Mechanisms of alpha-Synuclein Induced Synaptopathy in Parkinson's Disease. *Frontiers in neuroscience* 2018, *12*, 80.

- [35] Taylor, J. P., Hardy, J., Fischbeck, K. H., Toxic proteins in neurodegenerative disease. *Science* 2002, 296, 1991-1995.
- [36] Ross, C. A., Poirier, M. A., Protein aggregation and neurodegenerative disease. *Nature Medicine* 2004, 10 Suppl, S10-17.
- [37] Ross, C. A., Poirier, M. A., What is the role of protein aggregation in neurodegeneration? *Nature reviews Molecular cell biology* 2005, 6, 891.
- [38] Jellinger, K. A., Attems, J., Challenges of multimorbidity of the aging brain: a critical update. *Journal of neural transmission* 2015, 122, 505-521.
- [39] Whalley, L. J., Deary, I. J., Appleton, C. L., Starr, J. M., Cognitive reserve and the neurobiology of cognitive aging. *Ageing research reviews* 2004, 3, 369-382.
- [40] Stern, Y., Cognitive reserve. *Neuropsychologia* 2009, 47, 2015-2028.
- [41] Mayeux, R., Stern, Y., Epidemiology of Alzheimer disease. *Cold Spring Harbor perspectives in medicine* 2012, 2, a006239.
- [42] Alzheimer, A., Forstl, H., Levy, R., On certain peculiar diseases of old age. *History of psychiatry* 1991, 2, 71-101.
- [43] Alzheimer, A., *A Characteristic Disease of the Cerebral Cortex: Meeting of South-West Germany Psychiatrists Held in Tübingen on November 3rd and 4th, 1906*, 1907.
- [44] Murray, M. E., Graff-Radford, N. R., Ross, O. A., Petersen, R. C., *et al.*, Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* 2011, 10, 785-796.
- [45] Barnes, J., Dickerson, B. C., Frost, C., Jiskoot, L. C., *et al.*, Alzheimer's disease first symptoms are age dependent: Evidence from the NACC dataset. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2015, 11, 1349-1357.
- [46] Huang, Y., Mucke, L., Alzheimer mechanisms and therapeutic strategies. *Cell* 2012, 148, 1204-1222.
- [47] Graham, W. V., Bonito-Oliva, A., Sakmar, T. P., Update on Alzheimer's Disease Therapy and Prevention Strategies. *Annual review of medicine* 2017, 68.
- [48] Blennow, K., de Leon, M. J., Zetterberg, H., Alzheimer's disease. *Lancet* 2006, 368, 387-403.
- [49] Masliah, E., Mallory, M., Hansen, L., Richard, D., *et al.*, Synaptic and neuritic alterations during the progression of Alzheimer's disease. *Neuroscience Letters* 1994, 174, 67-72.
- [50] Wong, C. W., Quaranta, V., Glenner, G. G., Neuritic plaques and cerebrovascular amyloid in Alzheimer disease are antigenically related. *Proceedings of the National Academy of Sciences of the United States of America* 1985, 82, 8729-8732.
- [51] Kruger, U., Wang, Y., Kumar, S., Mandelkow, E. M., Autophagic degradation of tau in primary neurons and its enhancement by trehalose. *Neurobiol Aging* 2012, 33, 2291-2305.
- [52] Hardy, J. A., Higgins, G. A., Alzheimer's disease: the amyloid cascade hypothesis. *Science* 1992, 256, 184.

- [53] Sherrington, R., Rogaev, E., Liang, Y. a., Rogaeva, E., *et al.*, Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 1995, 375, 754.
- [54] Blennow, K., Bogdanovic, N., Alafuzoff, I., Ekman, R., Davidsson, P., Synaptic pathology in Alzheimer's disease: relation to severity of dementia, but not to senile plaques, neurofibrillary tangles, or the ApoE4 allele. *Journal of neural transmission* 1996, 103, 603-618.
- [55] Tiraboschi, P., Hansen, L. A., Masliah, E., Alford, M., *et al.*, Impact of APOE genotype on neuropathologic and neurochemical markers of Alzheimer disease. *Neurology* 2004, 62, 1977-1983.
- [56] van der Flier, W. M., Pijnenburg, Y. A., Fox, N. C., Scheltens, P., Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ϵ 4 allele. *The Lancet Neurology* 2011, 10, 280-288.
- [57] Price, J. L., McKeel Jr, D. W., Buckles, V. D., Roe, C. M., *et al.*, Neuropathology of nondemented aging: presumptive evidence for preclinical Alzheimer disease. *Neurobiology of aging* 2009, 30, 1026-1036.
- [58] Hardy, J., Selkoe, D. J., The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 2002, 297, 353-356.
- [59] Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., *et al.*, Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* 2002, 416, 535.
- [60] Lambert, M. P., Barlow, A., Chromy, B. A., Edwards, C., *et al.*, Diffusible, nonfibrillar ligands derived from A β 1-42 are potent central nervous system neurotoxins. *Proceedings of the National Academy of Sciences* 1998, 95, 6448-6453.
- [61] Benilova, I., Karran, E., De Strooper, B., The toxic A β oligomer and Alzheimer's disease: an emperor in need of clothes. *Nature neuroscience* 2012, 15, 349.
- [62] Karran, E., Mercken, M., De Strooper, B., The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nature reviews. Drug discovery* 2011, 10, 698-712.
- [63] Kimura, T., Whitcomb, D. J., Jo, J., Regan, P., *et al.*, Microtubule-associated protein tau is essential for long-term depression in the hippocampus. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2014, 369, 20130144.
- [64] Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., *et al.*, Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2016, 12, 292-323.
- [65] Fox, N. C., Crum, W. R., Scahill, R. I., Stevens, J. M., *et al.*, Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet* 2001, 358, 201-205.
- [66] Ganguli, M., Dodge, H. H., Shen, C., Pandav, R. S., DeKosky, S. T., Alzheimer disease and mortality: a 15-year epidemiological study. *Archives of neurology* 2005, 62, 779-784.
- [67] Kukull, W., Brenner, D., Speck, C., Nochlin, D., *et al.*, Causes of death associated with Alzheimer disease: variation by level of cognitive impairment before death. *Journal of the American Geriatrics Society* 1994, 42, 723-726.
- [68] Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., *et al.*, Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol* 2014, 13, 614-629.

- [69] McKhann, G., Drachman, D., Folstein, M., Katzman, R., *et al.*, Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984, 34, 939-939.
- [70] McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., *et al.*, The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2011, 7, 263-269.
- [71] Jack Jr, C. R., Lowe, V. J., Senjem, M. L., Weigand, S. D., *et al.*, 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnesic mild cognitive impairment. *Brain* 2008, 131, 665-680.
- [72] Chételat, G., Villemagne, V. L., Bourgeat, P., Pike, K. E., *et al.*, Relationship between atrophy and β -amyloid deposition in Alzheimer disease. *Annals of neurology* 2010, 67, 317-324.
- [73] Blennow, K., Mattsson, N., Scholl, M., Hansson, O., Zetterberg, H., Amyloid biomarkers in Alzheimer's disease. *Trends in pharmacological sciences* 2015, 36, 297-309.
- [74] Devanand, D., Mikhno, A., Pelton, G. H., Cuasay, K., *et al.*, Pittsburgh compound B (11C-PIB) and fluorodeoxyglucose (18 F-FDG) PET in patients with Alzheimer disease, mild cognitive impairment, and healthy controls. *Journal of geriatric psychiatry and neurology* 2010, 23, 185-198.
- [75] Hampel, H., Bürger, K., Teipel, S. J., Bokde, A. L., *et al.*, Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2008, 4, 38-48.
- [76] Frank, R. A., Galasko, D., Hampel, H., Hardy, J., *et al.*, Biological markers for therapeutic trials in Alzheimer's disease: proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease. *Neurobiology of aging* 2003, 24, 521-536.
- [77] Hampel, H., Toschi, N., Baldacci, F., Zetterberg, H., *et al.*, Alzheimer's disease biomarker-guided diagnostic workflow using the added value of six combined cerebrospinal fluid candidates: A β 1-42, total-tau, phosphorylated-tau, NFL, neurogranin, and YKL-40. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2018.
- [78] Sorbi, S., Hort, J., Erkinjuntti, T., Fladby, T., *et al.*, EFNS-ENS Guidelines on the diagnosis and management of disorders associated with dementia. *European Journal of Neurology* 2012, 19, 1159-1179.
- [79] Ascherio, A., Schwarzschild, M. A., The epidemiology of Parkinson's disease: risk factors and prevention. *Lancet Neurol* 2016, 15, 1257-1272.
- [80] Harold, D., Abraham, R., Hollingworth, P., Sims, R., *et al.*, Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature genetics* 2009, 41, 1088.
- [81] Reitz, C., Kunkle, B., Vardarajan, B., Kohli, M., *et al.*, Whole-Exome Sequencing Of Hispanic Early-Onset Alzheimer Disease Families Identifies Rare Variants In Multiple Alzheimer-Related Genes (S28.003). *Neurology* 2014, 82, S28.003.
- [82] Jellinger, K. A., Janetzky, B., Attems, J., Kienzl, E., Biomarkers for early diagnosis of Alzheimer disease: 'ALZheimer ASSociated gene'--a new blood biomarker? *Journal of cellular and molecular medicine* 2008, 12, 1094-1117.

- [83] Crary, J. F., Trojanowski, J. Q., Schneider, J. A., Abisambra, J. F., *et al.*, Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta neuropathologica* 2014, 128, 755-766.
- [84] Mischak, H., Allmaier, G., Apweiler, R., Attwood, T., *et al.*, Recommendations for biomarker identification and qualification in clinical proteomics. *Science translational medicine* 2010, 2, 46ps42.
- [85] Selkoe, D. J., Hardy, J., The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine* 2016, 8, 595-608.
- [86] Simonsen, A. H., Herukka, S. K., Andreasen, N., Baldeiras, I., *et al.*, Recommendations for CSF AD biomarkers in the diagnostic evaluation of dementia. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2017, 13, 274-284.
- [87] Trinh, J., Farrer, M., Advances in the genetics of Parkinson disease. *Nature reviews. Neurology* 2013, 9, 445-454.
- [88] Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., Tanzi, R. E., Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature genetics* 2007, 39, 17-23.
- [89] Rohn, T. T., Proteolytic cleavage of apolipoprotein e4 as the keystone for the heightened risk associated with Alzheimer's disease. *International journal of molecular sciences* 2013, 14, 14908-14922.
- [90] Gulbrandsen, A., Vethe, H., Farag, Y., Oveland, E., *et al.*, In-depth characterization of the cerebrospinal fluid (CSF) proteome displayed through the CSF proteome resource (CSF-PR). *Molecular & cellular proteomics* 2014, 13, 3152-3163.
- [91] Qian, W. J., Jacobs, J. M., Liu, T., Camp, D. G., 2nd, Smith, R. D., Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. *Molecular & cellular proteomics* 2006, 5, 1727-1744.
- [92] Zougman, A., Pilch, B., Podtelejnikov, A., Kiehnopf, M., *et al.*, Integrated analysis of the cerebrospinal fluid peptidome and proteome. *Journal of proteome research* 2007, 7, 386-399.
- [93] Yuan, X., Desiderio, D. M., Human cerebrospinal fluid peptidomics. *Journal of mass spectrometry* 2005, 40, 176-181.
- [94] Stark, M., Danielsson, O., Griffiths, W. J., Jornvall, H., Johansson, J., Peptide repertoire of human cerebrospinal fluid: novel proteolytic fragments of neuroendocrine proteins. *Journal of chromatography. B, Biomedical sciences and applications* 2001, 754, 357-367.
- [95] Brumback, R. A., *The cerebrospinal fluid*, Springer 1989, pp. 15-43.
- [96] Matsumae, M., Sato, O., Hirayama, A., Hayashi, N., *et al.*, Research into the physiology of cerebrospinal fluid reaches a new horizon: intimate exchange between cerebrospinal fluid and interstitial fluid may contribute to maintenance of homeostasis in the central nervous system. *Neurologia medico-chirurgica* 2016, 56, 416-441.
- [97] Duits, F. H., Martinez-Lage, P., Paquet, C., Engelborghs, S., *et al.*, Performance and complications of lumbar puncture in memory clinics: Results of the multicenter lumbar puncture feasibility study. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2016, 12, 154-163.
- [98] Kaplan, G., The psychogenic etiology of headache post lumbar puncture. *Psychosomatic medicine* 1967, 29, 376-379.

- [99] Selkoe, D. J., Alzheimer's disease: genes, proteins, and therapy. *Physiological reviews* 2001, 81, 741-766.
- [100] Portelius, E., Price, E., Brinkmalm, G., Stiteler, M., *et al.*, A novel pathway for amyloid precursor protein processing. *Neurobiol Aging* 2011, 32, 1090-1098.
- [101] Müller, U. C., Zheng, H., Physiological functions of APP family proteins. *Cold Spring Harbor perspectives in medicine* 2011, a006288.
- [102] Kaether, C., Skehel, P., Dotti, C. G., Axonal membrane proteins are transported in distinct carriers: a two-color video microscopy study in cultured hippocampal neurons. *Molecular biology of the cell* 2000, 11, 1213-1224.
- [103] Wiltfang, J., Esselmann, H., Bibl, M., Hüll, M., *et al.*, Amyloid β peptide ratio 42/40 but not A β 42 correlates with phospho-Tau in patients with low-and high-CSF A β 40 load. *Journal of neurochemistry* 2007, 101, 1053-1059.
- [104] Lewczuk, P., Lelental, N., Spitzer, P., Maler, J. M., Kornhuber, J., Amyloid-beta 42/40 cerebrospinal fluid concentration ratio in the diagnostics of Alzheimer's disease: validation of two novel assays. *Journal of Alzheimer's disease* 2015, 43, 183-191.
- [105] Mietelska-Porowska, A., Wasik, U., Goras, M., Filipek, A., Niewiadomska, G., Tau protein modifications and interactions: their role in function and dysfunction. *International journal of molecular sciences* 2014, 15, 4671-4713.
- [106] Seiberlich, V., Bauer, N. G., Schwarz, L., Ffrench-Constant, C., *et al.*, Downregulation of the microtubule associated protein tau impairs process outgrowth and myelin basic protein mRNA transport in oligodendrocytes. *Glia* 2015, 63, 1621-1635.
- [107] Majounie, E., Cross, W., Newsway, V., Dillman, A., *et al.*, Variation in tau isoform expression in different brain regions and disease states. *Neurobiology of aging* 2013, 34, 1922. e1927-1922. e1912.
- [108] Makrides, V., Shen, T. E., Bhatia, R., Smith, B. L., *et al.*, Microtubule-dependent Oligomerization of Tau IMPLICATIONS FOR PHYSIOLOGICAL TAU FUNCTION AND TAUOPATHIES. *Journal of Biological Chemistry* 2003, 278, 33298-33304.
- [109] Rosenberg, K. J., Ross, J. L., Feinstein, H. E., Feinstein, S. C., Israelachvili, J., Complementary dimerization of microtubule-associated tau protein: Implications for microtubule bundling and tau-mediated pathogenesis. *Proceedings of the National Academy of Sciences* 2008.
- [110] Mietelska-Porowska, A., Wasik, U., Goras, M., Filipek, A., Niewiadomska, G., Tau Protein Modifications and Interactions: Their Role in Function and Dysfunction. *International Journal of Molecular Sciences* 2014, 15, 4671-4713.
- [111] Meredith, J. E., Jr., Sankaranarayanan, S., Guss, V., Lanzetti, A. J., *et al.*, Characterization of novel CSF Tau and ptau biomarkers for Alzheimer's disease. *PLoS One* 2013, 8, e76523.
- [112] Kosik, K. S., Joachim, C. L., Selkoe, D. J., Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease. *Proceedings of the National Academy of Sciences* 1986, 83, 4044-4048.
- [113] Lee, V. M., Goedert, M., Trojanowski, J. Q., Neurodegenerative tauopathies. *Annual review of neuroscience* 2001, 24, 1121-1159.
- [114] Hernandez, F., Avila, J., Tauopathies. *Cellular and molecular life sciences* 2007, 64, 2219-2233.

- [115] Karch, C. M., Jeng, A. T., Goate, A. M., Extracellular Tau levels are influenced by variability in Tau that is associated with tauopathies. *Journal of Biological Chemistry* 2012, 287, 42751-42762.
- [116] Feinstein, H. E., Benbow, S. J., LaPointe, N. E., Patel, N., *et al.*, Oligomerization of the microtubule-associated protein tau is mediated by its N-terminal sequences: implications for normal and pathological tau action. *Journal of neurochemistry* 2016, 137, 939-954.
- [117] Smith, D. H., Meaney, D. F., Axonal damage in traumatic brain injury. *The neuroscientist* 2000, 6, 483-495.
- [118] Otto, M., Wiltfang, J., Tumani, H., Zerr, I., *et al.*, Elevated levels of tau-protein in cerebrospinal fluid of patients with Creutzfeldt-Jakob disease. *Neuroscience letters* 1997, 225, 210-212.
- [119] Öst, M., Nylen, K., Csajbok, L., Öhrfelt, A. O., *et al.*, Initial CSF total tau correlates with 1-year outcome in patients with traumatic brain injury. *Neurology* 2006, 67, 1600-1604.
- [120] Desai, A., Mitchison, T. J., Microtubule polymerization dynamics. *Annual review of cell and developmental biology* 1997, 13, 83-117.
- [121] Nogales, E., Structural insights into microtubule function. *Annual review of biochemistry* 2000, 69, 277-302.
- [122] Johnson, G. V., Stoothoff, W. H., Tau phosphorylation in neuronal cell function and dysfunction. *Journal of cell science* 2004, 117, 5721-5729.
- [123] Biernat, J., Mandelkow, E.-M., The development of cell processes induced by tau protein requires phosphorylation of serine 262 and 356 in the repeat domain and is inhibited by phosphorylation in the proline-rich domains. *Molecular biology of the cell* 1999, 10, 727-740.
- [124] Cho, J. H., Johnson, G. V., Primed phosphorylation of tau at Thr231 by glycogen synthase kinase 3 β (GSK3 β) plays a critical role in regulating tau's ability to bind and stabilize microtubules. *Journal of neurochemistry* 2004, 88, 349-358.
- [125] Uversky, V. N., Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Frontiers in aging neuroscience* 2015, 7, 18.
- [126] Janning, D., Igaev, M., Sündermann, F., Brühmann, J., *et al.*, Single-molecule tracking of tau reveals fast kiss-and-hop interaction with microtubules in living neurons. *Molecular biology of the cell* 2014, 25, 3541-3551.
- [127] Leroy, K., Brion, J.-P., Developmental expression and localization of glycogen synthase kinase-3 β in rat brain. *Journal of chemical neuroanatomy* 1999, 16, 279-293.
- [128] Ishiguro, K., Shiratsuchi, A., Sato, S., Omori, A., *et al.*, Glycogen synthase kinase 3 β is identical to tau protein kinase I generating several epitopes of paired helical filaments. *FEBS Letters* 1993, 325, 167-172.
- [129] Lovestone, S., Davis, D. R., Webster, M.-T., Kaech, S., *et al.*, Lithium reduces tau phosphorylation: effects in living cells and in neurons at therapeutic concentrations. *Biological psychiatry* 1999, 45, 995-1003.
- [130] Liu, S. J., Zhang, A. H., Li, H. L., Wang, Q., *et al.*, Overactivation of glycogen synthase kinase-3 by inhibition of phosphoinositol-3 kinase and protein kinase C leads to hyperphosphorylation of tau and impairment of spatial memory. *Journal of neurochemistry* 2003, 87, 1333-1344.

- [131] Spittaels, K., Van den Haute, C., Van Dorpe, J., Geerts, H., *et al.*, Glycogen synthase kinase-3 β phosphorylates protein tau and rescues the axonopathy in the central nervous system of human four-repeat tau transgenic mice. *Journal of Biological Chemistry* 2000, 275, 41340-41349.
- [132] Mi, K., Johnson, G. V., The role of tau phosphorylation in the pathogenesis of Alzheimer's disease. *Current Alzheimer Research* 2006, 3, 449-463.
- [133] Hanger, D. P., Anderton, B. H., Noble, W., Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. *Trends in molecular medicine* 2009, 15, 112-119.
- [134] Wang, Y., Mandelkow, E., Degradation of tau protein by autophagy and proteasomal pathways. *Biochemical Society transactions* 2012, 40, 644-652.
- [135] Poppek, D., Keck, S., Ermak, G., Jung, T., *et al.*, Phosphorylation inhibits turnover of the tau protein by the proteasome: influence of RCAN1 and oxidative stress. *Biochem Journal* 2006, 400, 511-520.
- [136] Jeganathan, S., von Bergen, M., Mandelkow, E. M., Mandelkow, E., The natively unfolded character of tau and its aggregation to Alzheimer-like paired helical filaments. *Biochemistry* 2008, 47, 10526-10539.
- [137] Buerger, K., Ewers, M., Pirttilä, T., Zinkowski, R., *et al.*, CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain* 2006, 129, 3035-3041.
- [138] Shay, J., *the Cytoskeleton*, Springer Science & Business Media 2013.
- [139] Liu, Q., Xie, F., Siedlak, S. L., Nunomura, A., *et al.*, Neurofilament proteins in neurodegenerative diseases. *Cellular and molecular life sciences* 2004, 61, 3057-3075.
- [140] Liu, Q., Xie, F., Siedlak, S. L., Nunomura, A., *et al.*, Neurofilament proteins in neurodegenerative diseases. *Cellular and Molecular Life Sciences CMLS* 2004, 61, 3057-3075.
- [141] Nixon, R. A., Shea, T. B., Dynamics of neuronal intermediate filaments: a developmental perspective. *Cell motility and the cytoskeleton* 1992, 22, 81-91.
- [142] Lee, M. K., Xu, Z., Wong, P. C., Cleveland, D. W., Neurofilaments are obligate heteropolymers in vivo. *The Journal of cell biology* 1993, 122, 1337-1350.
- [143] Zhou, W., Zhang, J., Ye, F., Xu, G., *et al.*, Plasma neurofilament light chain levels in Alzheimer's disease. *Neuroscience letters* 2017, 650, 60-64.
- [144] Lista, S., Toschi, N., Baldacci, F., Zetterberg, H., *et al.*, Diagnostic accuracy of CSF neurofilament light chain protein in the biomarker-guided classification system for Alzheimer's disease. *Neurochemistry international* 2017, 108, 355-360.
- [145] Gerendasy, D. D., Sutcliffe, J. G., RC3/neurogranin, a postsynaptic calpacitin for setting the response threshold to calcium influxes. *Molecular Neurobiology* 1997, 15, 131-163.
- [146] Huang, K. P., Huang, F. L., Jager, T., Li, J., *et al.*, Neurogranin/RC3 enhances long-term potentiation and learning by promoting calcium-mediated signaling. *Journal of Neuroscience* 2004, 24, 10660-10669.
- [147] Lista, S., Hampel, H., Synaptic degeneration and neurogranin in the pathophysiology of Alzheimer's disease. *Expert review of neurotherapeutics* 2017, 17, 47-57.

- [148] Standridge, J. B., Vicious cycles within the neuropathophysiologic mechanisms of Alzheimer's disease. *Current Alzheimer Research* 2006, 3, 95-107.
- [149] Masliah, E., Mallory, M., Hansen, L., DeTeresa, R., *et al.*, Synaptic and neuritic alterations during the progression of Alzheimer's disease. *Neuroscience letters* 1994, 174, 67-72.
- [150] Honer, W. G., Pathology of presynaptic proteins in Alzheimer's disease: more than simple loss of terminals. *Neurobiology of aging* 2003, 24, 1047-1062.
- [151] DeKosky, S. T., Scheff, S. W., Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Annals of neurology* 1990, 27, 457-464.
- [152] Kvartsberg, H., Duits, F. H., Ingelsson, M., Andreasen, N., *et al.*, Cerebrospinal fluid levels of the synaptic protein neurogranin correlates with cognitive decline in prodromal Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2015, 11, 1180-1190.
- [153] Kvartsberg, H., Portelius, E., Andreasson, U., Brinkmalm, G., *et al.*, Characterization of the postsynaptic protein neurogranin in paired cerebrospinal fluid and plasma samples from Alzheimer's disease patients and healthy controls. *Alzheimer's research & therapy* 2015, 7, 40.
- [154] Thorsell, A., Bjerke, M., Gobom, J., Brunhage, E., *et al.*, Neurogranin in cerebrospinal fluid as a marker of synaptic degeneration in Alzheimer's disease. *Brain research* 2010, 1362, 13-22.
- [155] Robinson, R., In the Pipeline-Mild Cognitive Impairment: Neurogranin Predicts Rate of Cognitive Decline in Individuals at Risk for Alzheimer's Disease. *Neurology Today* 2018, 18, 7-8.
- [156] Portelius, E., Olsson, B., Höglund, K., Cullen, N. C., *et al.*, Cerebrospinal fluid neurogranin concentration in neurodegeneration: relation to clinical phenotypes and neuropathology. *Acta Neuropathologica* 2018, 1-14.
- [157] Wellington, H., Paterson, R. W., Suarez-Gonzalez, A., Poole, T., *et al.*, CSF neurogranin or tau distinguish typical and atypical Alzheimer disease. *Ann Clin Transl Neurol* 2018, 5, 162-171.
- [158] Tharanathan, R. N., Kittur, F. S., Chitin—the undisputed biomolecule of great potential. 2003.
- [159] Hamid, R., Khan, M. A., Ahmad, M., Ahmad, M. M., *et al.*, Chitinases: an update. *Journal of pharmacy & bioallied sciences* 2013, 5, 21.
- [160] Chaplin, D. D., Overview of the immune response. *The journal of allergy and clinical immunology* 2010, 125, S3-23.
- [161] Sutherland, T. E., Maizels, R. M., Allen, J. E., Chitinases and chitinase-like proteins: potential therapeutic targets for the treatment of T-helper type 2 allergies. *Clinical and experimental allergy* 2009, 39, 943-955.
- [162] Reese, T. A., Liang, H. E., Tager, A. M., Luster, A. D., *et al.*, Chitin induces accumulation in tissue of innate immune cells associated with allergy. *Nature* 2007, 447, 92-96.
- [163] Ober, C., Chupp, G. L., The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. *Current opinion in allergy and clinical immunology* 2009, 9, 401-408.
- [164] Kazakova, M. H., Sarafian, V. S., YKL-40-a novel biomarker in clinical practice? *Folia medica* 2009, 51, 5.

- [165] Bonne-Barkay, D., Zagadailov, P., Zou, H., Niyonkuru, C., *et al.*, YKL-40 expression in traumatic brain injury: an initial analysis. *Journal of neurotrauma* 2010, 27, 1215-1223.
- [166] Bonne-Barkay, D., Bissel, S. J., Wang, G., Fish, K. N., *et al.*, YKL-40, a marker of simian immunodeficiency virus encephalitis, modulates the biological activity of basic fibroblast growth factor. *The American journal of pathology* 2008, 173, 130-143.
- [167] Bonne-Barkay, D., Wang, G., Starkey, A., Hamilton, R. L., Wiley, C. A., In vivo CHI3L1 (YKL-40) expression in astrocytes in acute and chronic neurological diseases. *Journal of neuroinflammation* 2010, 7, 34.
- [168] Baldacci, F., Lista, S., Cavedo, E., Bonuccelli, U., Hampel, H., Diagnostic function of the neuroinflammatory biomarker YKL-40 in Alzheimer's disease and other neurodegenerative diseases. *Expert review of proteomics* 2017, 14, 285-299.
- [169] Baldacci, F., Toschi, N., Lista, S., Zetterberg, H., *et al.*, Two-level diagnostic classification using cerebrospinal fluid YKL-40 in Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2017, 13, 993-1003.
- [170] Janelidze, S., Hertze, J., Zetterberg, H., Landqvist Waldö, M., *et al.*, Cerebrospinal fluid neurogranin and YKL-40 as biomarkers of Alzheimer's disease. *Annals of clinical and translational neurology* 2016, 3, 12-20.
- [171] Alcolea, D., Carmona-Iragui, M., Suárez-Calvet, M., Sánchez-Saudinós, M. B., *et al.*, Relationship between β -secretase, inflammation and core cerebrospinal fluid biomarkers for Alzheimer's disease. *Journal of Alzheimer's Disease* 2014, 42, 157-167.
- [172] Wennström, M., Surova, Y., Hall, S., Nilsson, C., *et al.*, The inflammatory marker YKL-40 is elevated in cerebrospinal fluid from patients with Alzheimer's but not Parkinson's disease or dementia with Lewy bodies. *PloS one* 2015, 10, e0135458.
- [173] Nesvizhskii, A. I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of proteomics* 2010, 73, 2092-2123.
- [174] Yates III, J. R., Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry* 1998, 33, 1-19.
- [175] Wu, C. C., MacCoss, M. J., Shotgun proteomics: tools for the analysis of complex biological systems. *Current opinion in molecular therapeutics* 2002, 4, 242-250.
- [176] Nesvizhskii, A. I., Roos, F. F., Grossmann, J., Vogelzang, M., *et al.*, Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Molecular & Cellular Proteomics* 2006, 5, 652-670.
- [177] Colinge, J., Cusin, I., Reffas, S., Mahé, E., *et al.*, Experiments in searching small proteins in unannotated large eukaryotic genomes. *Journal of proteome research* 2005, 4, 167-174.
- [178] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 2007, 389, 1017-1031.
- [179] Fenselau, C., A review of quantitative methods for proteomic studies. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 2007, 855, 14-20.
- [180] Filiou, M. D., Martins-de-Souza, D., Guest, P. C., Bahn, S., Turck, C. W., To label or not to label: applications of quantitative proteomics in neuroscience research. *Proteomics* 2012, 12, 736-747.

- [181] Urban, P. L., Quantitative mass spectrometry: an overview. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 2016, 374, 20150382.
- [182] Zhu, W., Smith, J. W., Huang, C.-M., Mass spectrometry-based label-free quantitative proteomics. *BioMed Research International* 2009, 2010, 840518.
- [183] Campbell, J., Rezai, T., Prakash, A., Krastins, B., *et al.*, Evaluation of absolute peptide quantitation strategies using selected reaction monitoring. *Proteomics* 2011, 11, 1148-1152.
- [184] Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., *et al.*, Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & Cellular Proteomics* 2014, 13, 907-917.
- [185] Wieling, J., LC-MS-MS experiences with internal standards. *Chromatographia* 2002, 55, S107-S113.
- [186] Harwood, M. D., Russell, M. R., Neuhoff, S., Warhurst, G., Rostami-Hodjegan, A., Lost in centrifugation: accounting for transporter protein losses in quantitative targeted absolute proteomics. *Drug metabolism and disposition: the biological fate of chemicals* 2014, 42, 1766-1772.
- [187] Slawski, M., Hussong, R., Tholey, A., Jakoby, T., *et al.*, Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching. *BMC Bioinformatics* 2012, 13, 291.
- [188] Lortie, M., Bark, S., Blantz, R., Hook, V., Detecting low-abundance vasoactive peptides in plasma: progress toward absolute quantitation using nano liquid chromatography-mass spectrometry. *Analytical biochemistry* 2009, 394, 164-170.
- [189] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., *et al.*, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, 26, 966-968.
- [190] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* 2012, 9, 555.
- [191] Gallien, S., Duriez, E., Crone, C., Kellmann, M., *et al.*, Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & cellular proteomics* 2012, 11, 1709-1723.
- [192] Pannee, J., Portelius, E., Oppermann, M., Atkins, A., *et al.*, A selected reaction monitoring (SRM)-based method for absolute quantification of Abeta38, Abeta40, and Abeta42 in cerebrospinal fluid of Alzheimer's disease patients and healthy controls. *Journal of Alzheimer's disease* 2013, 33, 1021-1032.
- [193] Rauniyar, N., Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *International journal of molecular sciences* 2015, 16, 28566-28581.
- [194] Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C., Liotta, L. A., Clinical proteomics: translating benchside promise into bedside reality. *Nature reviews Drug discovery* 2002, 1, 683.
- [195] Anderson, N. L., Anderson, N. G., Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 1998, 19, 1853-1861.
- [196] Tyers, M., Mann, M., From genomics to proteomics. *Nature* 2003, 422, 193.
- [197] Liotta, L. A., Davis, J. B., Couch, R. D., Fredolini, C., *et al.*, *Molecular Pathology (Second Edition)*, Elsevier 2018, pp. 183-203.

- [198] Collins, F. S., McKusick, V. A., Implications of the Human Genome Project for medical science. *Journal of the American Medical Association* 2001, 285, 540-544.
- [199] Collins, F. S., Morgan, M., Patrinos, A., The Human Genome Project: lessons from large-scale biology. *Science* 2003, 300, 286-290.
- [200] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., *et al.*, From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature biotechnology* 1996, 14, 61.
- [201] Schrader, M., Schulz-Knappe, P., Peptidomics technologies for human body fluids. *Trends in Biotechnology* 2001, 19, S55-60.
- [202] Soloviev, M., Finch, P., Peptidomics: bridging the gap between proteome and metabolome. *Proteomics* 2006, 6, 744-747.
- [203] Svensson, M., Skold, K., Svenningsson, P., Andren, P. E., Peptidomics-based discovery of novel neuropeptides. *Journal of proteome research* 2003, 2, 213-219.
- [204] Schulz-Knappe, P., Zucht, H. D., Heine, G., Jurgens, M., *et al.*, Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Combinatorial chemistry & high throughput screening* 2001, 4, 207-217.
- [205] Kastin, A., *Handbook of biologically active peptides*, Academic press 2013.
- [206] Tinoco, A. D., Saghatelian, A., Investigating endogenous peptides and peptidases using peptidomics. *Biochemistry* 2011, 50, 7447-7461.
- [207] Russell, C. L., Mitra, V., Hansson, K., Blennow, K., *et al.*, Comprehensive Quantitative Profiling of Tau and Phosphorylated Tau Peptides in Cerebrospinal Fluid by Mass Spectrometry Provides New Biomarker Candidates. *Journal of Alzheimer's disease* 2017, 55, 303-313.
- [208] Kuzyk, M. A., Hardie, D. B., Yang, J., Smith, D. S., *et al.*, A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *Journal of proteome research* 2010, 9, 5422-5437.
- [209] Salvi, G., De Los Rios, P., Vendruscolo, M., Effective interactions between chaotropic agents and proteins. *Proteins: Structure, Function, and Bioinformatics* 2005, 61, 492-499.
- [210] Macdonald, R. D., Khajehpour, M., Effects of the protein denaturant guanidinium chloride on aqueous hydrophobic contact-pair interactions. *Biophysical Chemistry* 2015, 196, 25-32.
- [211] Möglich, A., Krieger, F., Kiefhaber, T., Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *Journal of molecular biology* 2005, 345, 153-162.
- [212] Creighton, T. E., Experimental studies of protein folding and unfolding. *Progress in biophysics and molecular biology* 1979, 33, 231-297.
- [213] Wedemeyer, W. J., Welker, E., Narayan, M., Scheraga, H. A., Disulfide bonds and protein folding. *Biochemistry* 2000, 39, 4207-4216.
- [214] Sechi, S., Chait, B. T., Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical Chemistry* 1998, 70, 5150-5158.

- [215] Cleland, W. W., Dithiothreitol, a new protective reagent for SH groups. *Biochemistry* 1964, 3, 480-482.
- [216] Hale, J. E., Butler, J. P., Gelfanova, V., You, J.-S., Knierman, M. D., A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. *Analytical biochemistry* 2004, 333, 174-181.
- [217] Dmitrenko, O., Thorpe, C., Bach, R. D., Mechanism of SN2 Disulfide Bond Cleavage by Phosphorus Nucleophiles. Implications for Biochemical Disulfide Reducing Agents. *The Journal of Organic Chemistry* 2007, 72, 8298-8307.
- [218] Anson, M., The reactions of iodine and iodoacetamide with native egg albumin. *The Journal of general physiology* 1940, 23, 321.
- [219] Lane, L. C., A simple method for stabilizing protein-sulfhydryl groups during SDS-gel electrophoresis. *Analytical biochemistry* 1978, 86, 655-664.
- [220] Reiber, H., Cerebrospinal fluid--physiology, analysis and interpretation of protein patterns for diagnosis of neurological diseases. *Multiple Sclerosis* 1998, 4, 99-107.
- [221] Blennow, K., Cerebrospinal fluid protein biomarkers for Alzheimer's disease. *NeuroRx* 2004, 1, 213-225.
- [222] Zougman, A., Pilch, B., Podtelejnikov, A., Kiehnopf, M., *et al.*, Integrated analysis of the cerebrospinal fluid peptidome and proteome. *Journal of proteome research* 2008, 7, 386-399.
- [223] Chiasserini, D., van Weering, J. R., Piersma, S. R., Pham, T. V., *et al.*, Proteomic analysis of cerebrospinal fluid extracellular vesicles: a comprehensive dataset. *Journal of proteomics* 2014, 106, 191-204.
- [224] Zhang, Y., Guo, Z., Zou, L., Yang, Y., *et al.*, Data for a comprehensive map and functional annotation of the human cerebrospinal fluid proteome. *Data Brief* 2015, 3, 103-107.
- [225] Mohring, T., Kellmann, M., Jurgens, M., Schrader, M., Top-down identification of endogenous peptides up to 9 kDa in cerebrospinal fluid and brain tissue by nanoelectrospray quadrupole time-of-flight tandem mass spectrometry. *Journal of mass spectrometry* 2005, 40, 214-226.
- [226] Holtta, M., Minthon, L., Hansson, O., Holmen-Larsson, J., *et al.*, An integrated workflow for multiplex CSF proteomics and peptidomics-identification of candidate cerebrospinal fluid biomarkers of Alzheimer's disease. *Journal of proteome research* 2015, 14, 654-663.
- [227] Fricker, L. D., Lim, J., Pan, H., Che, F. Y., Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass spectrometry reviews* 2006, 25, 327-344.
- [228] Holtta, M., Zetterberg, H., Mirgorodskaya, E., Mattsson, N., *et al.*, Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* 2012, 7, e42555.
- [229] Hu, L., Li, X., Jiang, X., Zhou, H., *et al.*, Comprehensive peptidome analysis of mouse livers by size exclusion chromatography prefractionation and nanoLC- MS/MS identification. *Journal of proteome research* 2007, 6, 801-808.
- [230] Judd, R. C., *The protein protocols handbook*, Springer 2009, pp. 193-200.
- [231] Hölttä, M., Quantitative neuroproteomics for biomarker discovery in Alzheimer's disease. 2014.

- [232] Hölttä, M., Hansson, O., Andreasson, U., Hertze, J., *et al.*, Evaluating amyloid- β oligomers in cerebrospinal fluid as a biomarker for Alzheimer's disease. 2013.
- [233] Greening, D. W., Simpson, R. J., A centrifugal ultrafiltration strategy for isolating the low-molecular weight (≤ 25 K) component of human plasma proteome. *Journal of proteomics* 2010, 73, 637-648.
- [234] Tucholska, M., Scozzaro, S., Williams, D., Ackloo, S., *et al.*, Endogenous peptides from biophysical and biochemical fractionation of serum analyzed by matrix-assisted laser desorption/ionization and electrospray ionization hybrid quadrupole time-of-flight. *Analytical biochemistry* 2007, 370, 228-245.
- [235] Shores, K. S., Knapp, D. R., Assessment approach for evaluating high abundance protein depletion methods for cerebrospinal fluid (CSF) proteomic analysis. *Journal of proteome research* 2007, 6, 3739-3751.
- [236] Zheng, X., Baker, H., Hancock, W. S., Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. *Journal of chromatography. A* 2006, 1120, 173-184.
- [237] Sigdel, T. K., Ling, X. B., Lau, K. H., Li, L., *et al.*, Urinary peptidomic analysis identifies potential biomarkers for acute rejection of renal transplantation. *Clinical Proteomics* 2009, 5, 103.
- [238] Poole, C. F., Gunatilleka, A. D., Sethuraman, R., Contributions of theory to method development in solid-phase extraction. *Journal of chromatography. A* 2000, 885, 17-39.
- [239] Buszewski, B., Szultka, M., Past, present, and future of solid phase extraction: a review. *Critical Reviews in Analytical Chemistry* 2012, 42, 198-213.
- [240] Poole, C. F., New trends in solid-phase extraction. *Trends in analytical chemistry* 2003, 22, 362-373.
- [241] Thongboonkerd, V., Chutipongtanate, S., Kanlaya, R., Systematic evaluation of sample preparation methods for gel-based human urinary proteomics: quantity, quality, and variability. *Journal of proteome research* 2006, 5, 183-191.
- [242] Zhou, S., Bailey, M. J., Dunn, M. J., Preedy, V. R., Emery, P. W., A quantitative investigation into the losses of proteins at different stages of a two-dimensional gel electrophoresis procedure. *Proteomics* 2005, 5, 2739-2747.
- [243] 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. *Analytical Chemistry* 2003, 75, 663-670.
- [244] Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., *et al.*, The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* 2013, 10, 730.
- [245] Camerini, S., Mauri, P., The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *Journal of chromatography A* 2015, 1381, 1-12.
- [246] Shackleton, C., Clinical steroid mass spectrometry: a 45-year history culminating in HPLC-MS/MS becoming an essential tool for patient diagnosis. *The Journal of steroid biochemistry and molecular biology* 2010, 121, 481-490.
- [247] Sentell, K. B., Dorsey, J. G., Retention mechanisms in reversed-phase liquid chromatography. Stationary-phase bonding density and partitioning. *Analytical chemistry* 1989, 61, 930-934.

- [248] Shen, Y., Zhao, R., Belov, M. E., Conrads, T. P., *et al.*, Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Analytical chemistry* 2001, 73, 1766-1775.
- [249] Hao, Z., Xiao, B., Weng, N., Impact of column temperature and mobile phase components on selectivity of hydrophilic interaction chromatography (HILIC). *Journal of separation science* 2008, 31, 1449-1464.
- [250] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry* 2005, 77, 6426-6434.
- [251] Yang, F., Shen, Y., Camp, D. G., 2nd, Smith, R. D., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert review of proteomics* 2012, 9, 129-134.
- [252] Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., Yates III, J. R., Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* 2013, 113, 2343-2394.
- [253] Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr* 1990, 499, 177-196.
- [254] Yoshida, T., Peptide separation by hydrophilic-interaction chromatography: a review. *Journal of biochemical and biophysical methods* 2004, 60, 265-280.
- [255] Köcher, 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. *Analytical chemistry* 2011, 83, 2699-2704.
- [256] Cao, Z., Tang, H. Y., Wang, H., Liu, Q., Speicher, D. W., Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes. *Journal of proteome research* 2012, 11, 3090-3100.
- [257] Chiu, C. W., Chang, C. L., Chen, S. F., Evaluation of peptide fractionation strategies used in proteome analysis. *Journal of separation science* 2012, 35, 3293-3301.
- [258] Stasyk, T., Huber, L. A., Zooming in: fractionation strategies in proteomics. *Proteomics* 2004, 4, 3704-3716.
- [259] Lee, W.-C., Lee, K. H., Applications of affinity chromatography in proteomics. *Analytical biochemistry* 2004, 324, 1-10.
- [260] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *Journal of separation science* 2005, 28, 1694-1703.
- [261] Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., *et al.*, Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 2011, 11, 2019-2026.
- [262] Batth, T. S., Francavilla, C., Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *Journal of proteome research* 2014, 13, 6176-6186.
- [263] Dwivedi, R. C., Spicer, V., Harder, M., Antonovici, M., *et al.*, Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Analytical Chemistry* 2008, 80, 7036-7042.
- [264] Lawrence, R. T., Searle, B. C., Llovet, A., Villén, J., Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. *Nature methods* 2016, 13, 431.

- [265] Eliuk, S., Makarov, A., Evolution of orbitrap mass spectrometry instrumentation. 2015.
- [266] Rauniyar, N., Parallel reaction monitoring: a targeted experiment performed using high resolution and high mass accuracy mass spectrometry. *International journal of molecular sciences* 2015, *16*, 28566-28581.
- [267] Diamandis, E. P., Mass spectrometry as a diagnostic and a cancer biomarker discovery tool opportunities and potential limitations. *Molecular & Cellular Proteomics* 2004, *3*, 367-378.
- [268] Liao, H., Wu, J., Kuhn, E., Chin, W., *et al.*, Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis & Rheumatism* 2004, *50*, 3792-3803.
- [269] Pierrakos, C., Vincent, J.-L., Sepsis biomarkers: a review. *Critical care* 2010, *14*, R15.
- [270] Wang, L., Skotland, T., Berge, V., Sandvig, K., Llorente, A., Exosomal proteins as prostate cancer biomarkers in urine: from mass spectrometry discovery to immunoassay-based validation. *European Journal of Pharmaceutical Sciences* 2017, *98*, 80-85.
- [271] Ayrton, J., Dear, G., Leavens, W., Mallett, D., Plumb, R., The use of turbulent flow chromatography/mass spectrometry for the rapid, direct analysis of a novel pharmaceutical compound in plasma. *Rapid communications in mass spectrometry* 1997, *11*, 1953-1958.
- [272] Edge, T., Turbulent flow chromatography in bioanalysis. *Bioanalytical Separations* 2003, *4*, 91.
- [273] Couchman, L., Turbulent flow chromatography in bioanalysis: a review. *Biomedical chromatography* 2012, *26*, 892-905.
- [274] Van Deemter, J., Zuiderweg, F., Klinkenberg, A. v., Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science* 1956, *5*, 271-289.
- [275] Pretorius, V., Smuts, T., Turbulent Flow Chromatography. A New Approach to Faster Analysis. *Analytical Chemistry* 1966, *38*, 274-281.
- [276] Chassaing, C., Luckwell, J., Macrae, P., Saunders, K., *et al.*, Direct analysis of crude plasma samples by turbulent flow chromatography/tandem mass spectrometry. *Chromatographia* 2000, *53*, 122-130.
- [277] Feller, W. B., Hill, J. A., White, P. L., Abraham, J. M., Google Patents 2003.
- [278] Beckey, H.-D., *Field Ionization Mass Spectrometry: International Series of Monographs in Analytical Chemistry*, Elsevier 2017.
- [279] Brodbelt, J., Focus on Novel Instrumentation in Mass Spectrometry and Ion Mobility Spectrometry. *Journal of The American Society for Mass Spectrometry* 2018, *29*, 211-212.
- [280] Harrison, A. G., *Chemical ionization mass spectrometry*, Routledge 2018.
- [281] Pavlovich, M. J., Musselman, B., Hall, A. B., Direct analysis in real time—Mass spectrometry (DART-MS) in forensic and security applications. *Mass spectrometry reviews* 2018, *37*, 171-187.
- [282] Annesley, T. M., Ion suppression in mass spectrometry. *Clinical Chemistry* 2003, *49*, 1041-1044.
- [283] Al-Tarawneh, S. K., Border, M. B., Dibble, C. F., Bencharit, S., Defining salivary biomarkers using mass spectrometry-based proteomics: a systematic review. *OmicS: a journal of integrative biology* 2011, *15*, 353-361.

- [284] Gillette, M. A., Carr, S. A., Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature methods* 2013, 10, 28-34.
- [285] Mahboob, S., Mohamedali, A., Ahn, S. B., Schulz-Knappe, P., *et al.*, Is isolation of comprehensive human plasma peptidomes an achievable quest? *Journal of proteomics* 2015, 127, 300-309.
- [286] De Hoffmann, E., Mass spectrometry. *Kirk-Othmer Encyclopedia of Chemical Technology* 2000.
- [287] Awad, H., Khamis, M. M., El-Aneed, A., Mass spectrometry, review of the basics: ionization. *Applied Spectroscopy Reviews* 2015, 50, 158-175.
- [288] Han, X., Aslanian, A., Yates III, J. R., Mass spectrometry for proteomics. *Current opinion in chemical biology* 2008, 12, 483-490.
- [289] El-Aneed, A., Cohen, A., Banoub, J., Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Applied Spectroscopy Reviews* 2009, 44, 210-230.
- [290] Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., *et al.*, Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry* 2006, 78, 2113-2120.
- [291] Koppenaal, D. W., Barinaga, C. J., Denton, M. B., Sperline, R. P., *et al.*, ACS Publications 2005.
- [292] Kraj, A., Desiderio, D. M., Nibbering, N. M., *Mass spectrometry: instrumentation, interpretation, and applications*, John Wiley & Sons 2008.
- [293] de Hoffmann, E., Tandem mass spectrometry: a primer. *Journal of mass spectrometry* 1996, 31, 129-137.
- [294] Wilm, M., Mann, M., Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry* 1996, 68, 1-8.
- [295] Karas, M., Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* 1988, 60, 2299-2301.
- [296] Juhasz, P., Costello, C. E., Biemann, K., Matrix-assisted laser desorption ionization mass spectrometry with 2-(4-hydroxyphenylazo) benzoic acid matrix. *Journal of the American Society for Mass Spectrometry* 1993, 4, 399-409.
- [297] Stafford Jr, G., Kelley, P., Syka, J., Reynolds, W., Todd, J., Recent improvements in and analytical applications of advanced ion trap technology. *International Journal of Mass Spectrometry and Ion Processes* 1984, 60, 85-98.
- [298] Guan, S., Marshall, A. G., Stored waveform inverse Fourier transform axial excitation/ejection for quadrupole ion trap mass spectrometry. *Analytical Chemistry* 1993, 65, 1288-1294.
- [299] March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of mass spectrometry* 1997, 32, 351-369.
- [300] Hopfgartner, G., Varesio, E., Tschäppät, V., Grivet, C., *et al.*, Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *Journal of Mass Spectrometry* 2004, 39, 845-855.
- [301] March, R. E., Hughes, R. J., *Quadrupole storage mass spectrometry*, Wiley 1989.
- [302] Schwartz, J. C., Senko, M. W., Syka, J. E., A two-dimensional quadrupole ion trap mass spectrometer. *Journal of The American Society for Mass Spectrometry* 2002, 13, 659-669.

- [303] Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L., Park, B. K., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *Journal of Chromatography B* 2009, 877, 1229-1239.
- [304] Yost, R. A., Enke, C. G., Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Analytical Chemistry* 1979, 51, 1251-1264.
- [305] Dawson, P. H., *Quadrupole mass spectrometry and its applications*, Elsevier 2013.
- [306] Splendore, M., Londry, F., March, R., Morrison, R., *et al.*, A simulation study of ion kinetic energies during resonant excitation in a stretched ion trap. *International journal of mass spectrometry and ion processes* 1996, 156, 11-29.
- [307] Cody, R., Burnier, R., Freiser, B., Collision-induced dissociation with Fourier transform mass spectrometry. *Analytical chemistry* 1982, 54, 96-101.
- [308] Papayannopoulos, I. A., The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass spectrometry reviews* 1995, 14, 49-73.
- [309] Wells, J. M., McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Methods in enzymology* 2005, 402, 148-185.
- [310] DeSouza, L. V., Romaschin, A. D., Colgan, T. J., Siu, K. M., Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. *Analytical chemistry* 2009, 81, 3462-3470.
- [311] Barnidge, D. R., Dratz, E. A., Martin, T., Bonilla, L. E., *et al.*, Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Analytical chemistry* 2003, 75, 445-451.
- [312] Dass, C., *Fundamentals of contemporary mass spectrometry*, John Wiley & Sons 2007.
- [313] Petrović, M., Škrbić, B., Živančev, J., Ferrando-Climent, L., Barcelo, D., Determination of 81 pharmaceutical drugs by high performance liquid chromatography coupled to mass spectrometry with hybrid triple quadrupole-linear ion trap in different types of water in Serbia. *Science of the total environment* 2014, 468, 415-428.
- [314] Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry* 2000, 72, 1156-1162.
- [315] Hu, Q., Noll, R. J., Li, H., Makarov, A., *et al.*, The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry* 2005, 40, 430-443.
- [316] Eliuk, S., Makarov, A., Evolution of orbitrap mass spectrometry instrumentation. *Annual review of analytical chemistry (Palo Alto, Calif.)* 2015, 8, 61-80.
- [317] Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., *et al.*, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics* 2005, 4, 2010-2021.
- [318] Scigelova, M., Makarov, A., Orbitrap mass analyzer—overview and applications in proteomics. *Proteomics* 2006, 6, 16-21.
- [319] Perry, R. H., Cooks, R. G., Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* 2008, 27, 661-699.

- [320] Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods in enzymology* 1990, *193*, 886.
- [321] Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., *et al.*, Evaluation of HCD-and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & Cellular Proteomics* 2011, mcp. M111. 009910.
- [322] Sleno, L., Volmer, D. A., Ion activation methods for tandem mass spectrometry. *Journal of mass spectrometry* 2004, *39*, 1091-1112.
- [323] Olsen, J. V., Macek, B., Lange, O., Makarov, A., *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* 2007, *4*, 709.
- [324] Marzluff, E. M., Campbell, S., Rodgers, M., Beauchamp, J., Low-energy dissociation pathways of small deprotonated peptides in the gas phase. *Journal of the American Chemical Society* 1994, *116*, 7787-7796.
- [325] Tang, X. J., Boyd, R. K., Bertrand, M., An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid communications in mass spectrometry* 1992, *6*, 651-657.
- [326] Wysocki, V. H., Tsaprailis, G., Smith, L. L., Brechi, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry* 2000, *35*, 1399-1406.
- [327] Zubarev, R. A., Kelleher, N. L., McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* 1998, *120*, 3265-3266.
- [328] 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. *Proceedings of the National Academy of Sciences* 2004, *101*, 9528-9533.
- [329] Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., *et al.*, Electron capture dissociation for structural characterization of multiply charged protein cations. *Analytical chemistry* 2000, *72*, 563-573.
- [330] Good, D. M., Wirtala, M., McAlister, G. C., Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Molecular & Cellular Proteomics* 2007, *6*, 1942-1951.
- [331] Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., *et al.*, The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2006, *1764*, 1811-1822.
- [332] 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. *Journal of the American Society for Mass Spectrometry* 1996, *7*, 532-540.
- [333] Kalli, A., Smith, G. T., Sweredoski, M. J., Hess, S., Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers. *Journal of proteome research* 2013, *12*, 3071-3086.
- [334] Andrews, G. L., Dean, R. A., Hawkrigde, A. M., Muddiman, D. C., Improving proteome coverage on a LTQ-Orbitrap using design of experiments. *Journal of the American Society for Mass Spectrometry* 2011, *22*, 773-783.

- [335] Kalli, A., Hess, S., Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. *Proteomics* 2012, *12*, 21-31.
- [336] Kim, M.-S., Kandasamy, K., Chaerkady, R., Pandey, A., Assessment of resolution parameters for CID-based shotgun proteomic experiments on the LTQ-Orbitrap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 2010, *21*, 1606-1611.
- [337] Wong, C. C., Cociorva, D., Venable, J. D., Xu, T., Yates, J. R., Comparison of different signal thresholds on data dependent sampling in Orbitrap and LTQ mass spectrometry for the identification of peptides and proteins in complex mixtures. *Journal of the American Society for Mass Spectrometry* 2009, *20*, 1405-1414.
- [338] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology* 2006, *24*, 971-983.
- [339] Deutsch, E. W., File formats commonly used in mass spectrometry proteomics. *Molecular & cellular proteomics* 2012, *11*, 1612-1621.
- [340] Megger, D. A., Bracht, T., Meyer, H. E., Sitek, B., Label-free quantification in clinical proteomics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2013, *1834*, 1581-1590.
- [341] Bateman, K. P., Kellmann, M., Muenster, H., Papp, R., Taylor, L., Quantitative-qualitative data acquisition using a benchtop Orbitrap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 2009, *20*, 1441-1450.
- [342] Winiewski, J. R., Hein, M. Y., Cox, J., Mann, M., A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Molecular & cellular proteomics* 2014, mcp. M113. 037309.
- [343] Huttenhain, R., Malmstrom, J., Picotti, P., Aebersold, R., Perspectives of targeted mass spectrometry for protein biomarker verification. *Current opinion in chemical biology* 2009, *13*, 518-525.
- [344] Bantscheff, M., Lemeer, S., Savitski, M. M., Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* 2012, *404*, 939-965.
- [345] Cham, J. A., Bianco, L., Bessant, C., Free computational resources for designing selected reaction monitoring transitions. *Proteomics* 2010, *10*, 1106-1126.
- [346] Ong, S.-E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* 2005, *1*, 252-262.
- [347] Rauniyar, N., Yates, J. R., 3rd, Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research* 2014, *13*, 5293-5309.
- [348] Thompson, A., Schafer, J., Kuhn, K., Kienle, S., *et al.*, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry* 2003, *75*, 1895-1904.
- [349] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., *et al.*, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics* 2004, *3*, 1154-1169.

- [350] Xiang, F., Ye, H., Chen, R., Fu, Q., Li, L., N, N-dimethyl leucines as novel isobaric tandem mass tags for quantitative proteomics and peptidomics. *Analytical chemistry* 2010, 82, 2817-2825.
- [351] Sandberg, A., Branca, R. M., Lehtio, J., Forshed, J., Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference. *Journal of proteomics* 2014, 96, 133-144.
- [352] Boonen, K., De Haes, W., Van Houtven, J., Verdonck, R., *et al.*, *Peptidomics*, Springer 2018, pp. 141-159.
- [353] Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., *et al.*, Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews* 2012, 41, 3912-3928.
- [354] Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., *et al.*, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & Cellular Proteomics* 2011, 10, mcp. M111. 011015.
- [355] Steen, H., Küster, B., Mann, M., Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *Journal of mass spectrometry* 2001, 36, 782-790.
- [356] Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., *et al.*, Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry* 2006, 78, 2113-2120.
- [357] Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., *et al.*, Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics* 2011, 10, M111 011015.
- [358] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., *et al.*, Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 2011, 7, 280-292.
- [359] Henley, S. M., Bates, G. P., Tabrizi, S. J., Biomarkers for neurodegenerative diseases. *Current opinion in neurology* 2005, 18, 698-705.
- [360] Berven, F. S., Kroksveen, A. C., Berle, M., Rajalahti, T., *et al.*, Pre-analytical influence on the low molecular weight cerebrospinal fluid proteome. *Proteomics. Clinical applications* 2007, 1, 699-711.
- [361] Cox, B., Emili, A., Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature protocols* 2006, 1, 1872-1878.
- [362] Spivak, M., Weston, J., Bottou, L., Kall, L., Noble, W. S., Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. *Journal of proteome research* 2009, 8, 3737-3745.
- [363] Brosch, M., Yu, L., Hubbard, T., Choudhary, J., Accurate and sensitive peptide identification with Mascot Percolator. *Journal of proteome research* 2009, 8, 3176-3181.
- [364] Ma, B., Zhang, K., Hendrie, C., Liang, C., *et al.*, PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* 2003, 17, 2337-2342.

- [365] Zhang, J., Xin, L., Shan, B., Chen, W., *et al.*, PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics* 2012, *11*, M111 010587.
- [366] Zhang, J., Xin, L., Shan, B., Chen, W., *et al.*, PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics* 2011.
- [367] Shteynberg, D., Nesvizhskii, A. I., Moritz, R. L., Deutsch, E. W., Combining results of multiple search engines in proteomics. *Molecular & Cellular Proteomics* 2013, *12*, mcp. R113. 027797.
- [368] Kroksveen, A. C., Opsahl, J. A., Aye, T. T., Ulvik, R. J., Berven, F. S., Proteomics of human cerebrospinal fluid: discovery and verification of biomarker candidates in neurodegenerative diseases using quantitative proteomics. *Journal of proteomics* 2011, *74*, 371-388.
- [369] Hoglund, K., Fourier, A., Perret-Liaudet, A., Zetterberg, H., *et al.*, Alzheimer's disease--Recent biomarker developments in relation to updated diagnostic criteria. *Clinica chimica acta; international journal of clinical chemistry* 2015, *449*, 3-8.
- [370] Novakova, L., Vlckova, H., A review of current trends and advances in modern bio-analytical methods: chromatography and sample preparation. *Analytica chimica acta* 2009, *656*, 8-35.
- [371] Finoulst, I., Pinkse, M., Van Dongen, W., Verhaert, P., Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. *BioMed Research International* 2011, *2011*.
- [372] Oberhauser, C. J., Niggebrugge, A. E., Lachance, D., Takarewski, J. J., *et al.*, Turbulent-flow LC for LC-MS and LC-MS-MS bioanalysis. *LC GC North America* 2000, *18*, 716-724.
- [373] Shushan, B., A review of clinical diagnostic applications of liquid chromatography–tandem mass spectrometry. *Mass spectrometry reviews* 2010, *29*, 930-944.
- [374] Pan, J., Zhang, C., Zhang, Z., Li, G., Review of online coupling of sample preparation techniques with liquid chromatography. *Analytica chimica acta* 2014, *815*, 1-15.
- [375] Wu, J.-T., Zeng, H., Qian, M., Brogdon, B. L., Unger, S. E., Direct Plasma Sample Injection in Multiple-Component LC– MS– MS Assays for High-Throughput Pharmacokinetic Screening. *Analytical chemistry* 2000, *72*, 61-67.
- [376] Herman, J., The use of turbulent flow chromatography and the isocratic focusing effect to achieve on-line cleanup and concentration of neat biological samples for low-level metabolite analysis. *Rapid Communications in Mass Spectrometry* 2005, *19*, 696-700.
- [377] Fumes, B. H., Andrade, M. A., Franco, M. S., Lancas, F. M., On-line approaches for the determination of residues and contaminants in complex samples. *Journal of separation science* 2017, *40*, 183-202.
- [378] Herman, J. L., Edge, T., Majors, R. E., Theoretical concepts and applications of turbulent flow chromatography. *LC GC North America* 2012, *30*.
- [379] Song, Y., Song, Q., Li, J., Zheng, J., *et al.*, An integrated platform for directly widely-targeted quantitative analysis of feces part I: Platform configuration and method validation. *Journal of Chromatography A* 2016, *1454*, 58-66.
- [380] Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., *et al.*, Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *Journal of molecular biology* 1987, *195*, 659-685.

- [381] Meek, J. L., Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. *Proceedings of the National Academy of Sciences* 1980, 77, 1632-1636.
- [382] Eng, J. K., McCormack, A. L., Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 1994, 5, 976-989.
- [383] Falkner, J. A., Falkner, J. W., Yocum, A. K., Andrews, P. C., A spectral clustering approach to MS/MS identification of post-translational modifications. *Journal of proteome research* 2008, 7, 4614-4622.
- [384] Frank, A. M., Bandeira, N., Shen, Z., Tanner, S., *et al.*, Clustering millions of tandem mass spectra. *Journal of proteome research* 2007, 7, 113-122.
- [385] Beer, I., Barnea, E., Ziv, T., Admon, A., Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* 2004, 4, 950-960.
- [386] Wan, K. X., Vidavsky, I., Gross, M. L., Comparing similar spectra: from similarity index to spectral contrast angle. *Journal of the American Society for Mass Spectrometry* 2002, 13, 85-88.
- [387] Skillbäck, T., Mattsson, N., Hansson, K., Mirgorodskaya, E., *et al.*, A novel quantification-driven proteomic strategy identifies an endogenous peptide of pleiotrophin as a new biomarker of Alzheimer's disease. *Scientific Reports* 2017, 7, 13333.
- [388] González-Castillo, C., Ortuño-Sahagún, D., Guzmán-Brambila, C., Pallàs, M., Rojas-Mayorquín, A. E., Pleiotrophin as a central nervous system neuromodulator, evidences from the hippocampus. *Frontiers in cellular neuroscience* 2015, 8, 443.
- [389] Ryan, E., Shen, D., Wang, X., Structural studies reveal an important role for the pleiotrophin C-terminus in mediating interactions with chondroitin sulfate. *The FEBS journal* 2016, 283, 1488-1503.
- [390] Silos-Santiago, I., Yeh, H. J., Gurrieri, M. A., Guillerman, R. P., *et al.*, Localization of pleiotrophin and its mRNA in subpopulations of neurons and their corresponding axonal tracts suggests important roles in neural-glia interactions during development and in maturity. *Journal of neurobiology* 1996, 31, 283-296.
- [391] Yeh, H.-J., He, Y. Y., Xu, J., Hsu, C. Y., Deuel, T. F., Upregulation of pleiotrophin gene expression in developing microvasculature, macrophages, and astrocytes after acute ischemic brain injury. *Journal of Neuroscience* 1998, 18, 3699-3707.
- [392] Wanaka, A., Carroll, S. L., Milbrandt, J., Developmentally regulated expression of pleiotrophin, a novel heparin binding growth factor, in the nervous system of the rat. *Developmental brain research* 1993, 72, 133-144.
- [393] David, D. C., Layfield, R., Serpell, L., Narain, Y., *et al.*, Proteasomal degradation of tau protein. *Journal of neurochemistry* 2002, 83, 176-185.