Cerebrospinal Fluid Proteomics of Multiple Sclerosis Patients

Marcel Stoop

ISBN: 978-90-8559-918-0

Layout and printing: Optima Grafische Communicatie, Rotterdam, the Netherlands.

© M.P. Stoop. All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system of any nature, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission by the author.

Cerebrospinal Fluid Proteomics of Multiple Sclerosis Patients

Proteomics op liquor cerebrospinalis van multiple sclerosis patiënten

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 11 februari 2010 om 13:30 uur

door

Marcel Paulus Stoop

geboren te Rotterdam

PROMOTIECOMMISSIE

Promotor: Prof.dr. R.Q. Hintzen

Copromotor: Dr. T.M. Luider

Overige leden: Prof.dr. P.A.E. Sillevis Smitt
Prof.dr. J.D. Laman
Prof.dr. M.M. Breteler

Table of contents

1. Introduction: Multiple sclerosis-related biomarkers — 7
   Predictive Diagnostics and Personalized Treatment: Dream or Reality? (2009) Editor: O. Golubnitschaja, Chapter 18, page 355-373.
2. Quantitative proteomics and metabolomics analysis of normal human CSF samples — 27
   Manuscript submitted for publication.
3. The rate of false positive sequence matches of peptides profiled by MALDI MS and identified by MS/MS — 51
   Journal of Proteome Research (2008), 7: 4841-4847.
4. Multiple sclerosis related proteins identified in CSF by advanced mass spectrometry — 67
   Proteomics (2008), 8:1576-1585.
5. Quantitative matrix-assisted laser desorption ionization-fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple-sclerosis-related proteins — 87
   Journal of Proteome Research (2009), 8:1404-1414.
6. CSF differential proteomics of Relapsing Remitting and Primary Progressive Multiple Sclerosis — 113
   Manuscript submitted for publication.
7. Discussion — 131

   Summary — 155

   Samenvatting — 159

Acknowledgements / dankwoord — 163

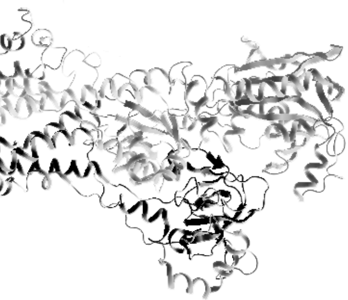
List of publications — 165

PhD portfolio — 167

Curriculum vitae — 171

List of abbreviations

2D-GE	2 dimensional gel electrophoresis
ACN	acetonitrile
AGC	automatic gain control
BBB	blood brain barrier
CIS	clinically isolated syndrome (of demyelination)
CNS	central nervous system
CSF	cerebrospinal fluid
CV	coefficient of variation
Da	dalton
DHB	2,5-dihydroxybenzoic acid
DTT	1,4-dithiothreitol
EDSS	expanded disability status scale
ELISA	enzyme linked immunosorbent assay
ESI	electrospray ionization
FTMS	fourier transform mass spectrometry
FT-ICR	fourier transform ion cyclotron resonance
GC	gas chromatography
HCCA	alpha-cyano-4-hydroxy cinamic acid
HCl	hydrochloric acid
HPLC	high-performance liquid chromatography
HUPO	human proteome organization
IEF	Isoelectric focussing
IgG	immunoglobulin gamma
IgK	immunoglobulin kappa
LC	liquid chromatography
LTQ	linear trap quadrupole
MALDI	matrix-assisted laser desorption/ionization
MeOH	methanol
MIDAS	MRM Initiated Detection and Sequencing
MRI	magnetic resonance imaging
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MScl	multiple sclerosis
n	number of samples/patients
NMR	nuclear magnetic resonance
OIND	other inflammatory neurological disease
OND	other neurological disease
PP (MScl)	primary progressive (multiple sclerosis)
ppm	parts-per-million
QTOF	quadrupole time-of-flight
rpm	rotations-per-minute
R^2	correlation coefficient
RR (MScl)	relapsing remitting (multiple sclerosis)
RSD	relative standard deviation
SD	standard deviation
SELDI	surface enhanced laser desorption/ionization
SRM	selected reaction monitoring
S/N	signal-to-noise
TFA	trifluoroacetic acid
TOF	time-of-flight
XC	cross correlation score



Chapter 1

Multiple sclerosis and the search for multiple sclerosis-related biomarkers

Marcel P. Stoop, Rogier Q. Hintzen, and Theo M. Luider

Predictive Diagnostics and Personalized Treatment: Dream or Reality? (2009) Editor: O. Golubnitschaja, Chapter 18, page 355-373.

Multiple sclerosis (MScl) is a highly heterogeneous disease of the central nervous system, and its pathology is characterized by a combination of factors such as inflammation, demyelination and axonal damage [1, 2]. Cerebrospinal fluid (CSF) is a relatively interesting body fluid in which to search for biomarkers and disease-associated proteins and peptides for MScl, due to its close proximity to disease processes [3, 4]. Although multiple sclerosis has been extensively researched, the cause of the disease remains elusive.

BACKGROUND

Even though MScl was first depicted in 1838, serious study of this demyelinating disease began in the latter part of the 19th century with the studies of Charcot [5]. He referred to the disease by a variety of different French names. After translation to other languages, a number of different names were used worldwide. Eventually the publication of *Multiple Sclerosis* in 1955 [6] allowed for a consensus to be reached, and this is the name the condition has been universally known as since then.

Currently MScl affects approximately one million people between ages 17 and 65 worldwide [7]. Although the number of people afflicted with the disease has increased steadily over time, a recent report indicates that the prevalence and incidence rates of MScl appear to have been stable over the past twenty years [8]. However, another recent report shows that the incidence of MScl in Canada still increases due to environmental effects [9]. The annual costs for patient care in MScl differs per patient based on numerous factors, disease severity perhaps foremost among them [10]. Other influential factors include relapse status, treatment type and gender [11]. MScl is far more common in women than in men, but men have a tendency for later disease onset, and have a worse prognosis. This supports evidence of gender-dependent factors in etiology and variability in phenotype [12].

Diagnosis of MScl has commonly been based on clinical evidence, supported by laboratory tools such as CSF testing and evoked potentials. In 1997 an International Panel revised the diagnostic criteria for MScl (since then known as the McDonald criteria), formally adopting MRI evidence in MScl diagnosis (Figure 1) [13]. Later a limited re-revision of the McDonald criteria were published. For establishing a MScl diagnosis, emphasis remains on objective clinical and paraclinical findings and depends on evidence of dissemination of lesions in time and space [14]. Quick diagnosis in an early stage of the disease is important as early intervention appears beneficial to long-term disease evolution [15, 16].

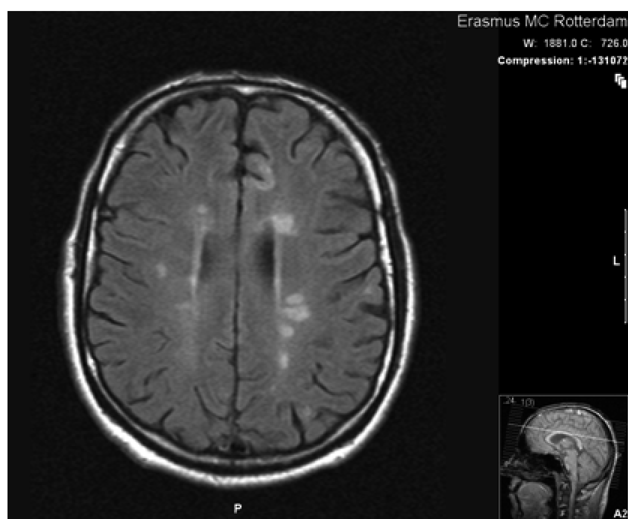


Figure 1. MRI image of a MScl patient, with MScl plaques highlighted.

Current management of relapsing remitting MScl is usually acute treatment of relapses with intravenous corticosteroids and maintenance disease-modifying therapy with first-line immunomodulatory drugs like glatiramer acetate or interferon-β [17], which is thought to relate to decrease of the pathologic inflammatory process through immune modulation and decreased trafficking of T-helper type 1 responses. This is an effective treatment, as interferon-β delays the time to diagnosis of definite MScl and reduced brain lesion burden in patients with clinical isolated syndromes of demyelination [18]. However, this is not effective in all patients. Recently, there has been a lot of interest in combining treatments, which would address more the heterogeneity of the disease process in MScl, with their several different pathophysiological processes occurring in parallel [2]. Natalizumab, a monoclonal antibody, is one of the most recent drugs that have been approved for treatment of MScl. Its effect is that it limits the trafficking of peripheral pro-inflammatory T cells into the central nervous system. Two trials, both reported on in 2006, examined the effect of treatment with natalizumab, both as a single therapy agent [19] as well as in combination with interferon-β [20]. Although the combined study showed significantly reduced relapse rates and fewer gadolinium enhanced lesions in patients treated with both natalizumab and interferon-β than compared to patients treated with only interferon-β. However, the relapse rates in these patients were not lower than in patients treated with only natalizumab, indicating there is no evidence for synergy between the two treatments. Other studies using glatiramer acetate in combination with the anticancer drug mitoxantrone do show enhanced efficacy of the combined therapy. Relapse activity was reduced [21] and blood testing showed that the mitoxantrone significantly attenuated the

proliferative response of B cells [22]. These results indicate a combination of the anti-inflammatory effects of mitoxantrone and a speculative neuroprotective effect of glatiramer acetate showed positive effects in these patients. Other combinations of therapeutic agents may also show beneficial effects in MScl treatment, and a significant number of these trials are currently underway.

MScl can be divided into two basic disease types: a relapsing type and a progressive type. In primary progressive MScl the disease at onset is progressive in nature and adverse effects and disability increase over time without periods of improvement. Relapsing remitting MScl is characterized by clearly defined relapses with full recovery in the period in between, marked by a lack of disease progression. After a period of relapsing remitting MScl, most patients progress into a secondary progressive MScl type of disease in which the affliction progressively worsens, with or without occasional relapses, minor remissions, and plateaus [23].

It takes a mean time of ten years for relapsing remitting MScl to progress to a secondary progressive phase that leads to irreversible disability [24, 25]. Primary progressive MScl, in which patients accumulate a progressive disability from the beginning, affect roughly 10-15% of MScl patients (Figure 2) [26]. Kurtzke´s expanded disability status scale (EDSS), in which higher scores indicated a more

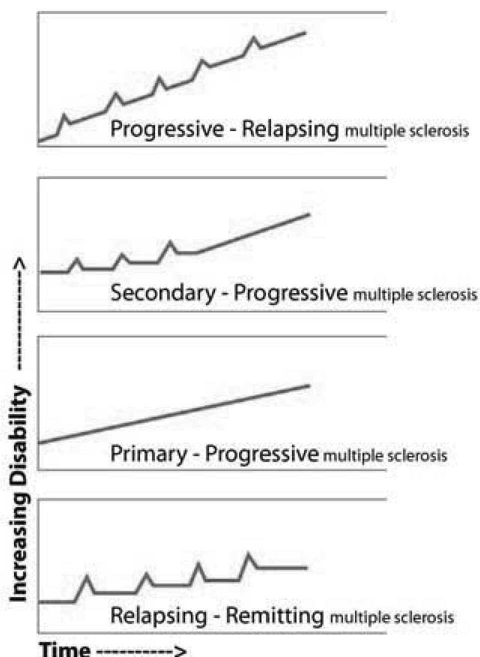


Figure 2. Graphical representation of disease progression of the different MScl disease types. (adapted from www.disaboom.com)

severe disability, is generally used to measure disability in MScl [27]. The EDSS scale ranges from zero to ten, with a score of six indicating that the patient is able to walk only with unilateral support and for no more than 100 meters. This score is considered to be an outcome that is representative of a poor evolution of the disease and the median time between onset of the disease and the assignment of an EDSS score of six is approximately twenty years [28, 29].

BIOLOGY

MScl is a central nervous system disease characterized by chronic inflammatory demyelination. Traditionally MScl has been considered an autoimmune disorder consisting of myelin autoreactive T cells that drive an inflammatory process, leading to myelin destruction (Figure 3) [30, 31]. However, MScl pathology is a great deal more complicated than that, which is not surprising given the heterogeneity observed in the clinical and phenotypical features of the disease [32]. The unique pathologic feature of MScl is the presence of multifocal demyelinated plaques scattered through the central nervous system [33]. Plaques are caused by lesions, which can be characterized as active or inactive. Active lesions are the regions in which demyelination is actively taking place [34]. Active MScl lesions are driven by inflammation yet evaluation of early lesions reveals four distinct pathological patterns [35, 36], which adds another aspect to the heterogeneity of the disease [37].

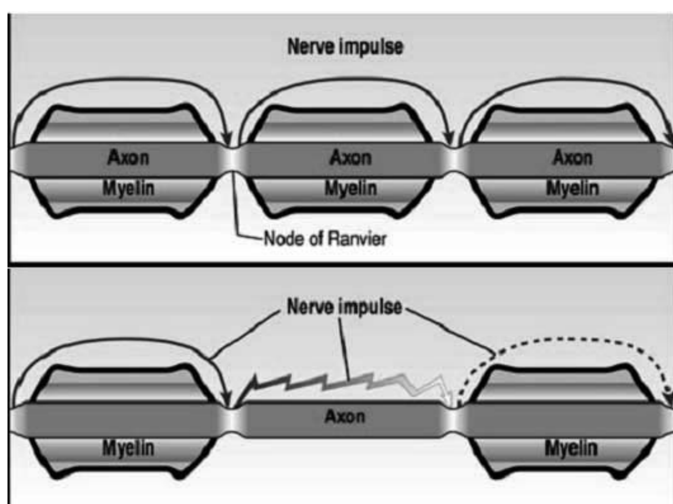


Figure 3. Signal transduction in nerve fibers in controls (top) and MScl (bottom). Due to demyelination in MScl, signal transduction is greatly reduced.
(adapted from www.multiplesclerosis.blogharbor.com)

The relapsing remitting phase of MScl seems to be characterized by the inflammatory and demyelinating component of the disease, whereas the progressive stage of the disease appears to be predominated by a neurodegenerative component, which leads to extensive neuroaxonal damage in the MScl brain. The accumulation of extensive axon damage that occurs from the early disease stages on is likely the cause of the loss of axons [38]. However, the timing and pathogenesis of axon destruction are not yet specifically defined, and may occur either in parallel with myelin destruction during the inflammatory phase or in a later, neurodegenerative, stage when the axon is demyelinated and more susceptible to damage [39-41].

BIOMARKERS

Since the inception of the term biomarker, it has often been used loosely, creating confusion and controversy about its definition. Attempts to clarify this have resulted in the following definition: A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes or pharmacological responses to a therapeutic intervention [2, 42, 43].

Due to the complexity and heterogeneity of MScl it is unlikely for a single biomarker to be found that either reflects all ongoing pathological processes or completely and accurately predicts the disease progression in individuals. Yet biomarkers are also potential intervention targets for new therapy options or may simply be used to elucidate different disease pathways. MScl biomarkers would thus likely be markers for specific processes ongoing in MScl patients, like for example immunological activation, which would likely fluctuate with the relapsing remitting course of the disease. Or markers for axonal damage, which may correlate better with the development of long-term disability and have high prognostic values [44].

Biomarker discoveries can be done in a variety of different biological matrices, like for example body fluids or tissue. In MScl, the brain lesions are of interest as they are the primary locations of disease activity. However, these can only be collected post-mortem, although histopathological analysis of MScl lesions has shown interesting results using immunoproteomics techniques [45]. For MScl biomarker studies in living subjects body fluids are a more suitable choice, and a number of options are available in this area. Urine is of potential interest as the method of collection is non-invasive, however, urinary tract and bladder problems are both common in more disabled MScl patients, which may negatively affect the results. Blood is relatively simple to collect; yet the collection is invasive and

the presence of an active clotting system complicates analysis. CSF collection is invasive, but the close proximity to the inflammatory lesions in the central nervous system may result in a better reflection of the relevant inflammatory processes in CSF compared to blood, and the absence of an active clotting system is preferable for most analysis techniques. Consequently, CSF is a promising biofluid in which to search for MScl related peptides and proteins [3, 4]. The total volume of CSF in humans is far lower than the total volume of blood, so this does limit the amount that can be sampled. CSF peptide profiling can be done without complicated pre-treatment steps, like high-abundant protein depletion, that are essential to serum peptide profiling, because the total protein load in CSF is lower than in serum, especially of high abundant proteins such as albumin [46, 47].

MScl biomarkers can be divided into two groups, inflammatory markers and neurodegenerative markers [48]. Inflammatory biomarkers correlate with MRI markers of disease activity [49]. Intermittent inflammatory activity in the central nervous system and periphery clearly takes place in MScl patients, when compared to controls [50, 51]. The pathological differences between relapsing remitting MScl and both primary progressive as well as secondary progressive MScl may be influenced by differences in inflammatory response, which have been previously described [52, 53]. Neurodegenerative biomarkers focus on neurobiological processes that are not necessarily unique to MScl. Myelin basic protein, which is released into the CSF during demyelination, can be used as an index of active demyelination [54]. The CSF levels of this protein are elevated in approximately 80% of MScl patients during clinical attacks. CSF levels of several markers for neuroaxonal loss are correlated with disability and are particularly elevated in progressive disease types [55, 56]. Levels of an acidic calcium binding protein (S100b) are elevated in CSF of MScl patients. The levels of this marker of astrocytosis and gliosis are particularly elevated in primary progressive MScl [57, 58]. There is also evidence for increased levels of neuroprotective factors in MScl, like, neuronal and glial adhesion factors and oligodendrocyte survival factors, indicating that remyelination and repair do take place in MScl [59-61].

Advanced unbiased discovery techniques like gene expression profiling [62] and proteomics are able to measure large numbers of samples in relatively short periods of time. These types of experiments yield large amounts of data which, when statistically analyzed, could play a significant role in the detection of novel candidate biomarkers. CSF proteomics is a rapidly developing area of MScl biomarker research that, although still relatively new, has shown some interesting discoveries.

Some of these proteomics studies were profiling studies to detect differential abundance of peptides and proteins in MScl patients and controls, whilst another

study explored the MScl CSF proteome to identify as many proteins as possible in MScl patients. Using samples of five MScl patients (4 relapsing remitting and one secondary progressive), Dumont *et al*. identified 65 proteins using a combination of two-dimensional gel electrophoresis and liquid chromatography tandem mass spectrometry [63]. Whilst this study revealed interesting proteins, with roughly a quarter of the identified proteins being identified in CSF for the first time, the set-up of this experiment precluded any comparison with control CSF. Several of the identified proteins were related to the inflammatory response system, confirming the CSF inflammatory activity in MScl. In subsequent experiments the same group used the same chromatography and mass spectrometry techniques to compare a pooled CSF sample of eight relapsing remitting MScl patients with a pooled CSF control sample of six cancer patients [64]. This study identified a total of 148 proteins in these two samples of which 80 were present in both samples. 24 proteins were present only in the pooled control sample and 44 proteins were only present in the pooled MScl sample. Although this study identified differences between the two samples, it is important to note that these samples were pooled samples; hence a single CSF sample from one patient may skew any comparison without the researchers knowledge. Also, due to the choice of control samples it is very well possible that the presented results are as much indicative of cancer CSF proteomics as they are of MScl CSF proteomics.

Using two-dimensional gel electrophoresis and peptide mass fingerprinting Hammack *et al*. identified four peptides to be present in a pooled CSF sample of three MScl patients that were not present in a pooled sample of three other patients with inflammation in the central nervous system, but no direct relation to MScl pathology was apparent [65]. Again, as with the previously mentioned study, it must be noted that these were pooled samples and the presented identified peptides may not be uniformly represented in CSF of MScl patients.

By using Surface Enhanced Laser Desorption/Ionization Time-Of-Flight mass spectrometry Irani *et al*. analysed 29 CSF samples of patients with MScl or clinically isolated syndrome (CIS) and compared the mass spectra to CSF samples of control patients (29 transverse myelitis and 50 HIV) [66]. In this comparison they found a unique protein that was 100% specific for MScl and CIS, although low levels of the protein were detected in some HIV patients. The protein was identified to be a cleavage product of cystatin C, which suggested that cleavage of cystatin C at that particular site may be MScl related. However, Del Boccio *et al*. later showed that the cleavage product of cystatin C is formed by degradation of the first eight N-terminal residues of cystatin C, and that this is not specific to CSF of MScl patients but rather the consequence of sample storage at –20°C instead of –80°C [67]. The conclusion of this study was that this cleavage product is not a

marker for MScl and that meticulous recording of sample storage conditions and uniform samples storage conditions across the entire population of samples used in a study is essential to acquiring reliable results.

A large study by O'Connor *et al.* details interesting similarities and differences in both serum and CSF of MScl patients and controls as well as some interesting background details on biofluid analysis [68]. Serum samples showed almost twice as many molecular components as CSF, whilst median variation for the proteins identified (i.e. CV's) was higher in CSF, which was thought to reflect a greater variation sample collection and a greater variation in clinical status of the patients in the CSF control group. In serum, the proteins that were differentially abundant in MScl samples came from a variety of biological processes, including inflammation, the complement cascade and acute phase reactants. Twice as many significant differences were found in the CSF comparison of MScl and control samples than were found in the serum comparison. This strongly supports the hypothesis that CSF is a better source than serum for initial identification of potential biomarkers for MScl. Many of the proteins identified in serum were also detected in CSF, but also several proteins produced in the central nervous system were detected in CSF. The differentially abundant proteins in serum can, like in CSF, be associated with a variety of biological processes.

CSF samples of 12 relapsing remitting MScl patients, 12 CIS patients and 24 neurological controls (headaches) were compared using two-dimensional (difference) gel electrophoresis [69]. In this study by Lehmensiek *et al.* protein spots that showed a more than two-fold difference between the patients and controls were selected for Matrix Assisted Laser Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry analysis. The results of this study showed one up-regulated and ten down-regulated proteins differing between MScl and controls and two up-regulated and eleven down-regulated proteins differing between CIS and controls. One of the proteins differing between CIS and controls was confirmed by immunoblot (Apolipoprotein A1).

A slightly different approach is shotgun proteomics. Here, instead of using a gel electrophoresis separation to detect differentials, a mass spectrometry method is employed to analyse differential abundance of proteins and peptides. In shotgun proteomics a CSF sample is digested enzymatically and measured by mass spectrometry. By using a very large number of samples and advanced statistics the peptide peaks that differ between the sample groups can be found. A second mass spectrometry step is then used to identify these differentially abundant peptides.

So, in summary, proteomics approaches to biomarker discovery for MScl have shown interesting results in recent years and these techniques have substantially added to the current knowledge of the biological changes characterizing this dis-

ease. As the field continues to expand, new options like new developed mass spectrometry machineries and software handling larger numbers of complex mass spactra have become available, enabling more in depth proteomics studies and these developments are essential for biomarker discovery research in MScl.

PERSPECTIVES

In the ever-evolving field of biomarker research new advancements are constantly made on many levels. New analysis techniques, machinery and software will substantially add to the researcher's demands. Yet sample collection and sample preparation remain problematic as standardization is still lacking in the clinic for proteomics experiments.

CSF moves from sites of formation, through the ventricles and subarachnoid spaces, to sites of reabsorption. Along the way, exchange with interstitial fluid in the surrounding neuropil continually modifies the composition of solutes in CSF [70], causing a variation in CSF composition for different sampling locations [71]. Because of this, standardization of CSF sampling in terms of location, sampling method and storage conditions are essential to obtain meaningful comparison between samples. Sample pretreatment must be identical in all ways before measurement in order to be able to make a good protein and peptide analysis of differences between samples, because if this is not the case then any found differences could easily be caused by other factors than the disease under investigation. Variation in sample treatment between different medical institutes could become a major problem in large CSF proteomics studies if for example CSF collection and storage conditions are not identical across the entire sample population. A similar issue exists for serum samples.

In many countries strict regulations apply to CSF sampling. In the Netherlands for example a medical reason must be present for sampling to take place. There is no medical reason to take CSF samples from healthy individuals, so there is no CSF available for healthy controls. Consequently samples from patients with other diseases must be chosen to serve as control samples. For this choice there are two basic options. 1- Select only samples from patients that are all afflicted with the same disease. 2- Select samples from patients with a variety of different diseases. The advantage of the first option is that differences will probably be more profound due to the homogeneity of the control group, while the disadvantage of this approach is that any found difference is just as likely to be caused by the disease in the control group as it is by the disease under investigation. The disadvantage of the second option is that differences between the disease under investigation and

the heterogeneous control group are likely to be less obvious and profound as in the first method, but the advantage is that if a differential marker is found, it is likely to be a robust finding.

One of the main issues is that currently a public CSF protein database is lacking. It is very unfortunate that all research groups are, due to the lack of a common used database, forced to rely on their own databases. Whilst research done in different databases is undeniably capable of yielding valuable results, it does make comparison between research results very difficult. A single database obtained by multiple centres will be an excellent first step towards integration of all available data into one, potentially a far clearer, picture of the current status of the CSF proteome will result.

High abundant proteins such as immunoglobulins and proteins related to the immune response observed in MScl hold a significant amount of interest in biomarker research. However, it is the low abundant proteins that are thought to be the real markers of changes in biology [72, 73]. To access these low abundant proteins the high abundant proteins first need to be removed from the sample [46]. This can be done using a variety of chromatography techniques, like for example affinity chromatography. Using this method, a number of proteins is (partially) removed from the sample, allowing for more in depth analysis of the low abundant proteins in the sample [74-76]. Although not 100% effective, this method does remove a large portion of the high abundant proteins, which is, even if not 100% effective, still beneficial for the results if the affinity separation is reproducible. Another option is isoelectric focussing in which the peptides are separated based on their physical characteristics [77, 78]. Large-scale application of these methods on CSF proteomics studies for MScl biomarkers could potentially open up a whole new range of peptides and proteins that could serve as markers for disease progression or elucidate certain pathological pathways in MScl.

New mass spectrometry methods could substantially add to the subject, especially if the new methods allow for quantitative analysis. A proteomics application for MALDI and nanoLC electrospray triple quadrupole mass spectrometers would be such a quantitative method. MALDI mass spectrometers have been shown to be able to quantitatively measure anti-HIV drugs in biological matrices [79]. Although this mass spectrometry technique seems to be better suited for small molecules [80, 81], which makes it not ideally suited for peptide profiling, it could be a very interesting tool for characterization of small peptide fragments generated by enzymatic digestion as well as natural occurring peptides, especially in light of its quantitative capacities. However, the constant development of new methods, technologies and apparatus that allows for new research options with each new development is extremely costly. But this does increase the researcher's capaci-

ties to discover and identify new biomarkers, as each new development potentially allows a new possibility in sample examination. Recent new developments in apparatus for example have allowed for additional mass spectrometry methods for protein quantitation [80, 81] and identification [82, 83].

As CSF sampling is not without risk standardization of the methods used is beneficial for both the reproducibility of the procedure, and subsequently its analysis results, as well as the safety of the patients. Ideally, successful CSF sampling should meet the following criteria: (1) obtain sufficient CSF on the first attempt, (2) occurs without trauma (i.e., CSF containing low levels of red blood cells per high powered field), (3) occurs with minimal discomfort to the patient during and after the procedure, and, (4) occurs without serious adverse events such as cerebral herniation [84]. Any CSF sample containing blood cells is contaminated by blood during the sampling procedure and consequently does not accurately portray the real biological situation in the CSF of patients [85]. Samples that contain more than a before hand defined number of red blood cells should be excluded from all proteomics CSF analysis, as this will influence the result in an unreliable way.

Proteomics analysis of paired samples (blood and CSF) is potentially a great asset to any study of diseases of the central nervous system [86], and this would also hold true for proteomics studies. If, for example well-defined CSF samples show high levels of serum proteins that can be correlated to matching serum samples, this would indicate that protein leakage in the blood brain barrier is taking place in that particular disease. Unfortunately such paired samples are often not obtainable in sufficient amounts, especially for control patients.

In mass spectrometry, analysis of individual mass spectra is possible with various software packages, but these packages lack the ability to compare large groups of samples. Statistical analysis of peptide profiles in proteomics biomarker studies requires fast, user-friendly software for high-throughput data analysis for the comparison of sample groups of controls and diseased patients [87]. The development of software applications for these types of analysis is currently ongoing and promising studies on the subject have been published, like for example a software application that allows for comparison of large numbers (n > 100) of raw Fourier Transform mass spectra to pinpoint differentially abundant peptide peaks [88]. In addition, recently developed commercial packages such as Sieve (version 2) (ThermoFisher Scientific) and Progenesis LC-MS (Nonlinear Dynamics) are also available for these analyses.

After identification of potential biomarker candidates the next step would be to validate the elevated or decreased abundance of these markers using different analysis methods or a new set of samples, to prevent any sample-based or technical bias, because validation is needed to determine the clinical value of

new biomarkers [89]. Most prospective proteomics biomarkers are low abundant proteins [72, 73], often no specific assays are developed for these proteins, which complicates the biomarker validation process, indicating the need for more validation methods.

CONCLUSIONS

The rapidly growing field of CSF proteomics in MScl biomarker discovery has shown to be of substantial added value to the field. This is not merely so for only MScl biomarker research, but also for other diseases like for example Guillain-Barré syndrome [90] and Alzheimer's disease [91]. Although various issues still need to be addressed in CSF proteomics (e.g. standardization of CSF collection and storage, the absence of relative and absolute values of protein concentrations of large numbers of CSF proteins) it is obvious that mass spectrometry techniques add an additional perspective to biomarker discovery in MS. The robustness of high-throughput mass spectrometry methods capable of measuring larger numbers of samples (n > 100) with high mass accuracy and clear quantitative elements is of great benefit in finding MScl related proteins. It allows for a great perspective on the biological processes on the central level of the genomics, proteomics and metabolomics cascade in MScl.

ACKNOWLEDGEMENT

Author MPS was sponsored by the Top Institute Pharma (grant D4-102-1).

REFERENCES

1. Compston, A., et al., *Multiple sclerosis.* Lancet, 2002. **359**: p. 1221-31.
2. Bielekova, B., et al., *Development of biomarkers in multiple sclerosis.* Brain, 2004. **127**: p. 1463-78.
3. Romeo, M.J., et al., *CSF proteome: a protein repository for potential biomarker identification.* Expert Rev Proteomics, 2005. **2**: p. 57-70.
4. Yuan, X., et al., *Proteomics analysis of prefractionated human lumbar cerebrospinal fluid.* Proteomics, 2005. **5**: p. 541-50.
5. Charcot, J.M., *Leçons sur les Maladies du Systeme Nerveux faites a la Salpétière*, in *Progres Medicale*, A. Delahaye, Lecrosnier, E., Editor. 1872: Paris. p. 1-495.
6. McAlpine, D., Compston, N. D., Lumsden C. E., *Multiple Sclerosis*. 1955, Edinburgh: Livingstone.
7. Kantarci, O., et al., *Epidemiology and natural history of multiple sclerosis: new insights.* Curr Opin Neurol, 2006. **19**: p. 248-54.
8. Mayr, W.T., et al., *Incidence and prevalence of multiple sclerosis in Olmsted County, Minnesota, 1985-2000.* Neurology, 2003. **61**: p. 1373-7.
9. Ebers, G.C., *Environmental factors and multiple sclerosis.* Lancet Neurol, 2008. **7**: p. 268-77.
10. Kobelt, G., et al., *Costs and quality of life in multiple sclerosis in The Netherlands.* Eur J Health Econ, 2006. **7 Suppl 2**: p. S55-64.
11. Tyas, D., et al., *The distribution of the cost of multiple sclerosis in the UK: how do costs vary by illness severity?* Value Health, 2007. **10**: p. 386-9.
12. Kantarci, O.H., et al., *Natural history of multiple sclerosis.* Neurol Clin, 2005. **23**: p. 17-38, v.
13. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis.* Ann Neurol, 2001. **50**: p. 121-7.
14. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria".* Ann Neurol, 2005. **58**: p. 840-6.
15. Tintore, M., *Early MS treatment.* Int MS J, 2007. **14**: p. 5-10.
16. Tintore, M., *Rationale for early intervention with immunomodulatory treatments.* J Neurol, 2008. **255 Suppl 1**: p. 37-43.
17. Gold, R., *Combination therapies in multiple sclerosis.* J Neurol, 2008. **255 Suppl 1**: p. 51-60.
18. Lam, S., et al., *Interferon-beta(1b) for the treatment of multiple sclerosis.* Expert Opin Drug Metab Toxicol, 2008. **4**: p. 1111-1117.
19. Polman, C.H., et al., *A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis.* N Engl J Med, 2006. **354**: p. 899-910.
20. Rudick, R.A., et al., *Natalizumab plus interferon beta-1a for relapsing multiple sclerosis.* N Engl J Med, 2006. **354**: p. 911-23.
21. Ramtahal, J., et al., *Sequential maintenance treatment with glatiramer acetate after mitoxantrone is safe and can limit exposure to immunosuppression in very active, relapsing remitting multiple sclerosis.* J Neurol, 2006. **253**: p. 1160-4.
22. Bar-Or, A., Gibbs, E., Niino, M., Aziz, T., Atalab, S., Shi, F., Campagnolo, D., Jalili, F., Rhodes, S., Yamashita, T., Fan, B., Freedman, M., Panitch, H., Arnold, D., Vollmer, T., Oger, J., *Trackin In Vivo Modulation in MS Patients Treated with Glatiramer Acetate (GA) Alone, or with GA*

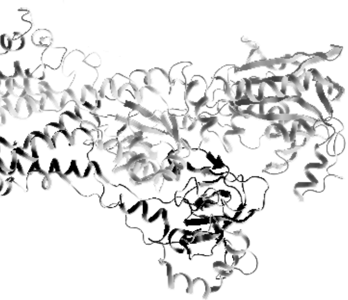
preceded by Mitoxantrone, Provides Novel Insights into the Mode of Action of Therapeutic Strategies That Combine Immune Suppression and Immune Deviation. Neurology, 2007. **68 (suppl 1)**: p. A18.

23. Compston, A., Confavreux, *The natural history of multiple sclerosis*, in *McAlpine's Multiple Sclerosis*. 2005, Churchill Livingstone Elsevier: Philadelphia. p. 183-272.
24. Rudick, R., et al., *Clinical outcomes assessment in multiple sclerosis.* Ann Neurol, 1996. **40**: p. 469-79.
25. Confavreux, C., et al., *Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients.* Brain, 1980. **103**: p. 281-300.
26. Confavreux, C., et al., *Relapses and progression of disability in multiple sclerosis*. N Engl J Med, 2000. **343**: p. 1430-8.
27. Kurtzke, J.F., *Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS).* Neurology, 1983. **33**: p. 1444-52.
28. Confavreux, C., et al., *Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process.* Brain, 2003. **126**: p. 770-82.
29. Weinshenker, B.G., et al., *The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course.* Brain, 1989. **112 (Pt 6)**: p. 1419-28.
30. Pittock, S.J., et al., *The pathology of MS: new insights and potential clinical applications.* Neurologist, 2007. **13**: p. 45-56.
31. McFarland, H.F., et al., *Multiple sclerosis: a complicated picture of autoimmunity.* Nat Immunol, 2007. **8**: p. 913-9.
32. Bielekova, B., et al., *MRI as a marker for disease heterogeneity in multiple sclerosis*. Neurology, 2005. **65**: p. 1071-6.
33. Lumsden, C., *The neuropathology of multiple sclerosis*, in *Handbook of Clincal Neurology*. 1970, Elsevier: New York. p. 217-309.
34. Lucchinetti, C.F., et al., *Multiple sclerosis: lessons from neuropathology.* Semin Neurol, 1998. **18**: p. 337-49.
35. Lucchinetti, C., et al., *Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination.* Ann Neurol, 2000. **47**: p. 707-17.
36. Ludwin, S.K., *The pathogenesis of multiple sclerosis: relating human pathology to experimental studies.* J Neuropathol Exp Neurol, 2006. **65**: p. 305-18.
37. Wingerchuk, D.M., et al., *Comparative immunopathogenesis of acute disseminated encephalomyelitis, neuromyelitis optica, and multiple sclerosis.* Curr Opin Neurol, 2007. **20**: p. 343-50.
38. Bruck, W., et al., *Inflammation and degeneration in multiple sclerosis.* Neurol Sci, 2003. **24 Suppl 5**: p. S265-7.
39. Ferguson, B., et al., *Axonal damage in acute multiple sclerosis lesions.* Brain, 1997. **120 (Pt 3)**: p. 393-9.
40. Kornek, B., et al., *Multiple sclerosis and chronic autoimmune encephalomyelitis: a comparative quantitative study of axonal injury in active, inactive, and remyelinated lesions.* Am J Pathol, 2000. **157**: p. 267-76.
41. Kuhlmann, T., et al., *Acute axonal damage in multiple sclerosis is most extensive in early disease stages and decreases over time.* Brain, 2002. **125**: p. 2202-12.
42. Lesko, L.J., et al., *Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies.* Annu Rev Pharmacol Toxicol, 2001. **41**: p. 347-66.

43. Rolan, P., et al., *Use of biomarkers from drug discovery through clinical practice: report of the Ninth European Federation of Pharmaceutical Sciences Conference on Optimizing Drug Development.* Clin Pharmacol Ther, 2003. **73**: p. 284-91.
44. van Waesberghe, J.H., et al., *Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability.* Ann Neurol, 1999. **46**: p. 747-54.
45. Han, M.H., et al., *Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets.* Nature, 2008. **451**: p. 1076-81.
46. Ramstrom, M., et al., *Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry.* J Proteome Res, 2005. **4**: p. 410-6.
47. You, J.S., et al., *The impact of blood contamination on the proteome of cerebrospinal fluid.* Proteomics, 2005. **5**: p. 290-6.
48. Giovannoni, G., *Multiple sclerosis cerebrospinal fluid biomarkers.* Dis Markers, 2006. **22**: p. 187-96.
49. Giovannoni, G., et al., *Correlating immunological and magnetic resonance imaging markers of disease activity in multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1998. **64 Suppl 1**: p. S31-6.
50. Gironi, M., et al., *Immunological markers in multiple sclerosis.* Neurol Sci, 2000. **21**: p. S871-5.
51. Uccelli, A., et al., *Biological markers of the inflammatory phase of multiple sclerosis.* Neurol Sci, 2003. **24 Suppl 5**: p. S271-4.
52. Thompson, A.J., et al., *Primary progressive multiple sclerosis.* Brain, 1997. **120 (Pt 6)**: p. 1085-96.
53. Weiner, L.P., et al., *Immunological aspects of secondary progressive multiple sclerosis.* Mult Scler, 2002. **8**: p. 83-4.
54. Cohen, S.R., et al., *A diagnostic index of active demyelination: myelin basic protein in cerebrospinal fluid.* Ann Neurol, 1980. **8**: p. 25-31.
55. Petzold, A., et al., *Axonal damage accumulates in the progressive phase of multiple sclerosis: three year follow up study.* J Neurol Neurosurg Psychiatry, 2005. **76**: p. 206-11.
56. Semra, Y.K., et al., *Heightened intrathecal release of axonal cytoskeletal proteins in multiple sclerosis is associated with progressive disease and clinical disability.* J Neuroimmunol, 2002. **122**: p. 132-9.
57. Michetti, F., et al., *The S-100 antigen in cerebrospinal fluid as a possible index of cell injury in the nervous system.* J Neurol Sci, 1980. **44**: p. 259-63.
58. Petzold, A., et al., *Markers for different glial cell responses in multiple sclerosis: clinical and pathological correlations.* Brain, 2002. **125**: p. 1462-73.
59. Massaro, A.R., *Are there indicators of remyelination in blood or CSF of multiple sclerosis patients?* Mult Scler, 1998. **4**: p. 228-31.
60. Massaro, A.R., et al., *N-CAM in cerebrospinal fluid: a marker of synaptic remodelling after acute phases of multiple sclerosis?* Ital J Neurol Sci, 1987. **Suppl 6**: p. 85-8.
61. Massaro, A.R., et al., *Cerebrospinal-fluid ciliary neurotrophic factor in neurological patients.* Eur Neurol, 1997. **37**: p. 243-6.
62. Mycko, M.P., et al., *cDNA microarray analysis in multiple sclerosis lesions: detection of genes associated with disease activity.* Brain, 2003. **126**: p. 1048-57.
63. Dumont, D., et al., *Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients.* Proteomics, 2004. **4**: p. 2117-24.

64. Noben, J.P., et al., *Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography, and mass spectrometry.* J Proteome Res, 2006. **5**: p. 1647-57.
65. Hammack, B.N., et al., *Proteomic analysis of multiple sclerosis cerebrospinal fluid.* Mult Scler, 2004. **10**: p. 245-60.
66. Irani, D.N., et al., *Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis.* Ann Neurol, 2006. **59**: p. 237-47.
67. Del Boccio, P., et al., *Cleavage of cystatin C is not associated with multiple sclerosis.* Ann Neurol, 2007. **62**: p. 201-4; discussion 205.
68. O'Connor, K.C., et al., *Comprehensive phenotyping in multiple sclerosis: discovery based proteomics and the current understanding of putative biomarkers.* Dis Markers, 2006. **22**: p. 213-25.
69. Lehmensiek, V., et al., *Cerebrospinal fluid proteome profile in multiple sclerosis.* Mult Scler, 2007. **13**: p. 840-9.
70. Shen, D.D., et al., *Principles and applicability of CSF sampling for the assessment of CNS drug delivery and pharmacodynamics.* Adv Drug Deliv Rev, 2004. **56**: p. 1825-57.
71. Finehout, E.J., et al., *Towards two-dimensional electrophoresis mapping of the cerebrospinal fluid proteome from a single individual.* Electrophoresis, 2004. **25**: p. 2564-75.
72. Burtis, C.A., Ashwood, E. R., *Tietz Fundamentals of Clinical Chemistry* 5th ed. 2001, Philadelphia: W.B. Saunders Company.
73. Gygi, S.P., et al., *Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.* Proc Natl Acad Sci U S A, 2000. **97**: p. 9390-5.
74. Martosella, J., et al., *Multi-component immunoaffinity subtraction and reversed-phase chromatography of human serum.* Methods Mol Biol, 2008. **425**: p. 27-39.
75. Ramstrom, M., et al., *Development of affinity columns for the removal of high-abundant proteins in cerebrospinal fluid.* Biotechnol Appl Biochem, 2008.
76. Dekker, L.J., et al., *Depletion of high-abundance proteins from serum by immunoaffinity chromatography: a MALDI-FT-MS study.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **847**: p. 65-9.
77. Stalder, D., et al., *Evaluation of reproducibility of protein identification results after multidimensional human serum protein separation.* Proteomics, 2008. **8**: p. 414-24.
78. Storms, H.F., et al., *Optimization of protein identification from digests as analyzed by capillary isoelectric focusing-mass spectrometry.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **824**: p. 189-200.
79. van Kampen, J.J., et al., *Quantitative analysis of antiretroviral drugs in lysates of peripheral blood mononuclear cells using MALDI-triple quadrupole mass spectrometry.* Anal Chem, 2008. **80**: p. 4969-75.
80. Corr, J.J., et al., *Design considerations for high speed quantitative mass spectrometry with MALDI ionization.* J Am Soc Mass Spectrom, 2006. **17**: p. 1129-41.
81. Gobey, J., et al., *Characterization and performance of MALDI on a triple quadrupole mass spectrometer for analysis and quantification of small molecules.* Anal Chem, 2005. **77**: p. 5643-54.
82. Hardman, M., et al., *Interfacing the orbitrap mass analyzer to an electrospray ion source.* Anal Chem, 2003. **75**: p. 1699-705.
83. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* J Mass Spectrom, 2005. **40**: p. 430-43.

84. Straus, S.E., et al., *How do I perform a lumbar puncture and analyze the results to diagnose bacterial meningitis?* Jama, 2006. **296**: p. 2012-22.
85. Pan, S., et al., *A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry.* Proteomics, 2007. **7**: p. 469-73.
86. Maimone, D., et al., *Cytokine levels in the cerebrospinal fluid and serum of patients with multiple sclerosis*. J Neuroimmunol, 1991. **32**: p. 67-74.
87. Titulaer, M.K., et al., *A database application for pre-processing, storage and comparison of mass spectra derived from patients and controls.* BMC Bioinformatics, 2006. **7**: p. 403.
88. Titulaer, M.K., et al., *A software application for comparing large numbers of high resolution MALDI-FTICR MS spectra demonstrated by searching candidate biomarkers for glioma blood vessel formation.* BMC Bioinformatics, 2008. **9**: p. 133.
89. Kohn, E.C., et al., *Proteomics as a tool for biomarker discovery.* Dis Markers, 2007. **23**: p. 411-7.
90. Jin, T., et al., *Proteomic identification of potential protein markers in cerebrospinal fluid of GBS patients.* Eur J Neurol, 2007. **14**: p. 563-8.
91. Finehout, E.J., et al., *Cerebrospinal fluid proteomic biomarkers for Alzheimer’s disease.* Ann Neurol, 2007. **61**: p. 120-9.



Chapter 2

Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples

Marcel P. Stoop#, Leon Coulier#, Therese Rosenling, Shanna Shi, Agnieszka M. Smolinska, Lutgarde Buydens, Kirsten Ampt, Christoph Stingl, Adrie Dane, Bas Muilwijk, Peter A. E. Sillevis Smitt, Rogier Q. Hintzen, Rainer Bischoff, Sybren S. Wijmenga, Thomas Hankemeier, Alain J. van Gool, and Theo M. Luider.

# Authors contributed equally to the study.

Manuscript submitted for publication.

ABSTRACT

The analysis of cerebrospinal fluid (CSF) is employed in biomarker discovery studies for various neurodegenerative central nervous system disorders. However, little is known about variation of CSF proteins and metabolites between patients without neurological disorders, a baseline for a large number of CSF compounds appears to be lacking. To analyze the variation in CSF protein and metabolite abundances in a number of well-defined individual samples of patients undergoing routine, non-neurological, surgical procedures we determined the variation of various proteins and metabolites by multiple analytical platforms.

A total of 130 common proteins were assessed for biological variations between individuals by ESI-Orbitrap. A large spread in inter-individual variation was observed (RSDs ranged from 18% to 148%), for both high abundant and low abundant proteins. Technical variation was between 15% and 30% for all 130 proteins. Metabolomics analysis was performed by means of GC-MS and NMR and amino acids were specifically analyzed by LC-MS/MS, resulting in the detection of more than 100 metabolites. Interestingly, the variation in the metabolome appears to be much more limited compared to the proteome, as the observed RSDs ranged from 16% to 63%. Technical variation was below 20% for almost all metabolites.

Consequently, an understanding of the biological variation of proteins and metabolites in CSF of neurologically normal individuals appears to be essential for reliable interpretation of biomarker discovery studies for central nervous system disorders, because such results may be influenced by natural inter-individual variations. Therefore proteins and metabolites with high variation between individuals ought to be assessed with caution as candidate biomarkers because at least part of the difference observed between the diseased individuals and the controls will not be caused by the disease, but rather by the natural biological variation between individuals.

INTRODUCTION

The analysis of cerebrospinal fluid (CSF) is indispensable in the diagnosis and understanding of various neurodegenerative central nervous system (CNS) disorders [1-3]. CSF is a fluid that has different functions, such as the protection of the brain to forces from outside, transport of biological substances and excretion of toxic and waste substances. It is in close contact with the extracellular fluid of the brain. Therefore, the composition of CSF can reflect biological processes of the brain [4]. By characterization of the proteome and metabolome of CSF better insight in, for example, the pathogenesis of CNS disorders may be achieved; as for many of these disorders the aetiology is still unclear.

CSF is produced in the ventricles of the brain and in the subarachnoidal spaces. Humans normally produce around 500 mL of CSF each day, and the total volume of CSF at a given time is approximately 150 mL. CSF reflects the composition of blood plasma although the concentrations of most proteins and metabolites in CSF are lower. However, individual proteins and metabolites can act differently. Active transport from blood and secretion from the brain contribute to the specific composition of CSF. This change in composition can be disturbed in neurological disorders [5, 6]. Since CNS specific proteins and metabolites are typically low in abundance compared to blood, this change in composition is more likely to be found in CSF and if the disease markers do not cross the blood-brain-barrier (BBB), then CSF is the only viable source. CSF might therefore be an excellent source for biomarker discovery for CNS disorders, following the hypothesis that neurological diseases induce alterations in CSF protein and metabolite levels.

Analysis of metabolites in CSF has been common practice in clinical chemistry for decades to analyse biomarkers for neurodegenerative and neuro-oncological disease diagnosis. The approaches used are either targeted analysis of one or a few metabolites using specific analytical methods [7], or metabolite profiling of CSF using NMR [8]. Metabolomics includes the analysis of metabolites in biofluids by NMR or MS-based approaches, i.e. LC-MS or GC-MS. Several metabolite profiling studies were carried out on CSF using NMR, some of which were published only recently [9, 10]. Surprisingly, very few metabolomics studies using MS-based methods have been performed on CSF yet [11, 12]. One of the reasons is the fact that the human CSF metabolome has not been characterized very well yet. Many CSF metabolites remain unidentified and for those that have been identified there is not much known about normal concentration ranges. A systematic categorization of the CSF metabolome is necessary and expected to be beneficial for future biomarker discoveries. Recently Wishart *et al.* made a good start in exploring the human CSF metabolome. Computer-aided literature survey resulted in 308 detectable metabolites in human CSF [13].

The CSF proteome has been characterized to a much larger extent than the CSF metabolome and is currently topic of investigations in several research groups worldwide. Recently, studies have been published with numerous identities and quantities of CSF proteins. Pan and co-workers were able to identify 2.594 proteins in well-characterized pooled human CSF samples using strict proteomics criteria with a combination of LTQ-FT and MALDI TOF/TOF equipment [14]. They were also able to quantify several proteins using a targeted LC MALDI TOF/TOF approach [15]. Hu and co-workers have studied the intra- and inter-individual variation in human CSF, and found large variations in protein concentrations in six patients by means of 2D-gel electrophoresis [16], focussing mainly on the variations within individuals at two different time-points. Although only a limited number of proteins was analyzed, the variation between the time-points was profound, exceeding 200% for seven proteins.

Unique CSF biomarkers may contribute to a deeper understanding of the mechanisms of CNS disorders. However, for this assumption to come true, there are still challenges ahead. Even though CSF is not as complex as blood (almost missing the cellular part and the clotting system present in blood), it is expected to consist of thousands of organic- and non-organic salts, sugars, lipids and proteins. A large part of the CSF consists of a few high abundant metabolites and proteins, which hamper, if no precautions are undertaken, the identification and quantification of metabolites and proteins that occur in lower amounts. The analysis of the CSF metabolome is complicated due to the diverse chemical nature of metabolites and the lower concentration of metabolites compared to blood. Analytical method development is still required as it is not possible to identify the entire range of CSF metabolites with one single analytical method. Though in proteome research efforts have been made to quantify proteins, metabolomics studies up to now do not provide quantitative information or only give information for the most abundant metabolites.

Another challenge is the sample amount obtained by lumbar puncture to collect CSF. Lumbar puncture is an invasive method that is not performed as frequently as blood sampling. However, often after the analysis of various clinical parameters only a limited amount of CSF sample is available for biomarker discovery. Metabolomics studies are hampered by limited CSF sample amount. Therefore analytical methods are required that are suitable to handle relative small sample volumes.

The main objective of this study was firstly to analyze the variation in CSF protein and metabolite abundances in a number of well-defined individual samples by multiple analytical platforms. Secondly, the goal was to integrate metabolomics and proteomics and to present biological variations in metabolite and protein abundances and compare these with technical variations with the currently employed

analytical methods. The results will facilitate and increase the application of CSF for future biomarker discovery studies in the field of neurodegenerative diseases and neuro-oncology.

EXPERIMENTAL PROCEDURES

CSF sampling

CSF samples were obtained by lumbar puncture in the Erasmus University Medical Centre (Rotterdam, the Netherlands). An experienced medical doctor selected ten samples, which were taken from patients receiving spinal anaesthesia prior to non-neurological surgery. These subjects had no neurological diseases, were not using any medication and were considered to have neurologically normal CSF. Immediately after sampling, the CSF samples were centrifuged (10 minutes at 3.000 rpm) to discard cellular elements. The samples were subsequently used for routine CSF diagnostics. This included quantification of total protein concentration by routine clinical chemistry measurements and quantification of the cell count (< 5 white blood cells per mL). The remaining volume of the samples was aliquoted and stored at –80°C immediately after centrifugation. As a standard procedure the samples were checked for blood contamination, and any sample in which hemoglobin or apolipoprotein B100 peptide was significantly identified by nanoLC-Orbitrap MS was excluded from the study.

For pooling of the samples (n=10), the originally obtained samples were thawed on ice and 0.75 mL from each of the samples was joined, resulting in a 7.5 mL pooled CSF sample. This pooled CSF sample was vortexed for 30 seconds and then subdivided into 75 portions of 100 µL in sterile cryogenic vials (Nalgene Nunc Int., Rochester, NY, USA). The portions were immediately frozen at –80°C. The characteristics of the pooled sample are described in Table 1. One of the samples was completely finished by pooling, so for the measurements of the individual patients only nine CSF samples were used.

The CSF samples used in the experimental sample set were selected by an experienced neurologist and taken from patients undergoing tests for clinical diagnosis. These samples, taken from multiple sclerosis and headache patients were

Table 1: Details on the pooled sample, including age, gender and protein concentration.

Gender	Male 8; Female 2
Mean age (years)	51 (SD = 14)
Total protein concentration (g/L)	0.4 (SD = 0.1)
Glucose concentration (mmol/L)	3.3 (SD = 0.3)

subjected to the same, strict post-sampling procedure as the samples mentioned previously. In these samples no significant difference in protein concentration between the two groups was observed, so there was no leakage in the blood-CSF barrier. The Medical Ethical Committee of the Erasmus University Medical Centre in Rotterdam, The Netherlands, approved the study protocol and all study participants gave written consent.

Proteomics

Sample preparation for nanoLC-Orbitrap MS and MALDI-FT-ICR MS
For measurement of proteins in CSF, samples were enzymatically digested with trypsin to obtain peptides. An amount of 50 µL Rapigest (Waters, USA) in 50 mM ammonium bicarbonate and 1 µL 100 mM DTT was added to 50 µL CSF. The mixture was heated at 60°C for 30 minutes, upon which it was cooled down to room temperature in approximately 20 minutes. Iodoacetamide (5 µL of 0.3 M solution) was added and this mixture was left for 30 minutes in dark at room temperature. Trypsin was added (10 µL, 0.1 mg/mL) and all samples, processed in one batch, were incubated overnight at 37°C. To stop digestion, 2 µL of a 50% TFA/50% water solution was added. The sample was then incubated for 45 minutes at 37°C.

NanoLC-Orbitrap MS analysis
These measurements were carried out on a Ultimate 3000 nano LC system (Dionex, Germering, Germany) online coupled to a hybrid linear ion trap / Orbitrap MS (LTQ Orbitrap XL; Thermo Fisher Scientific, Germany). Five µL digest were loaded on to a C18 trap column (C18 PepMap, 300µm ID x 5mm, 5µm particle size, 100 Å pore size; Dionex, The Netherlands) and desalted for 10 minutes using a flow rate of 20 µL /min 0.1% TFA. Then the trap column was switched online with the analytical column (PepMap C18, 75 µm ID x 150 mm, 3 µm particle and 100 Å pore size; Dionex, The Netherlands) and peptides were eluted with following binary gradient: 0% - 25% solvent B in 120 min and 25% - 50% solvent B in further 60 minutes, where solvent A consist of 2% acetonitrile and 0.1% formic in water and solvent B consists of 80% acetonitrile and 0.08% formic acid in water. Column flow rate was set to 300 nL/min. For MS detection a data dependent acquisition method was used: high resolution survey scan from 400 – 1800 Th. was performed in the Orbitrap (value of target of automatic gain control AGC 10^6, resolution 30,000 at 400 m/z; lock mass was set to 445.120025 u (protonated $(Si(CH_3)_2O)_6$) [17]). Based on this survey scan the 5 most intensive ions were consecutively isolated (AGC target set to 10^4 ions) and fragmented by collision-activated dissociation (CAD) applying 35% normalized collision energy in the linear ion trap. After precursors

were selected for MS/MS, they were excluded for further MS/MS spectra for 3 minutes. The MS/MS identifications were obtained using in the Bioworks 3.2 (peak picking by Extract_msn, default settings) software package (Thermo Fisher Scientific, Germany), and its SEQUEST feature, using the HUPO criteria, with XC scores of 1.8, 2.2 and 3.75 for single, double and triple charged ions respectively in the SwissProt-database (version 56.0, human taxonomy (20069 entries)). Carboxymethylation of Cysteine (+57.021 u) as fixed and oxidation of Methionine (+15.996 u) as variable modifications and tryptic cleavage were considered. The number of allowed missed cleavages was 2, the mass tolerance for precursor ions was 10 ppm and for fragment ions 0.5 Da. The cut-off for mass differences with the theoretical mass of the identified peptides was set at 2 ppm.

The Orbitrap data was subsequently analysed using the Progenesis LC-MS software package (version 2.5, Nonlinear Dynamics, United Kingdom), in which the LC runs were aligned and the biological variation between the samples was calculated to assess variation between individuals in this data set. A S/N > 4 and the presence of at least 3 isotope peaks per peptide were used as a minimum threshold for quantitation. Variation was assessed by comparing the area-under-the-curve of all peptides of a protein. The mean area-under-the-curve of all peptides of a protein was compared between the individuals, and the relative standard deviation (RSD) of this value was considered to be the inter-individual variation (listed as RSD (in percentages) in the supplementary material). Technical variation was assessed by performing the same comparison on the five replicas of the pooled sample.

MALDI-FT-ICR MS analysis

The CSF samples were handled according to the same protocol we reported previously [18], in which the samples were tryptically digested and desalted using C18 material. Using a 2,5-dihydroxybenzoic acid matrix the samples were all measured manually on an APEX IV Qe 9.6 Tesla MALDI-FT-ICR mass spectrometer (Bruker Daltonics, USA), using a multishot accumulation as recommended by Mize *et al.*, Moyer *et al.*, and O'Connor *et al.* [19-21]. External mass calibration was applied using a quadratic equation. Quantitative MALDI-FT-ICR has previously been applied to quantify HIV-1 protease inhibitors in cell lysates [22] as well as peptides in CSF [18], indicating that quantitative MALDI-FT-ICR methods are readily applicable, which is due to the fact that variation in peak height in MALDI-FT-ICR mass spectrometry is much more reproducible than in, for example, MALDI-TOF mass spectrometry. The sum of the height of 14 omnipresent albumin peaks of each sample was then compared to albumin concentrations obtained by routine clinical chemistry measurements. Standard deviations of the peak height of the albumin peaks were between 9-16% for all 14 albumin peaks.

Biological variation in an experimental setting

To test the results on the variation of protein abundances found in the nine individual CSF samples in an experimental setting, an identical experiment was performed on a larger set of samples. A total of 36 CSF samples, taken from patients with either multiple sclerosis or headaches, was used. It must be noted that these samples, especially those of the multiple sclerosis patients, originate from people suffering from neurological problems. Hence the variation in protein abundance, like for example immunoglobulin levels, which are known to be elevated in neuro-inflammatory diseases such as multiple sclerosis [23-25], is potentially far more extensive than in the nine well-defined individuals measured previously.

Metabolomics

GC-MS analysis

Human CSF samples (60 μL) were deproteinized by adding 250 μL methanol and subsequently centrifuged for 10 min at 10000 rpm. The supernatant was dried under N_2 followed by derivatization with methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) in pyridine similar to Koek *et al.* [26]. During the different steps in the sample work-up, i.e. prior to deproteinization, derivatization and injection, different (deuterated) internal standards were added at a level of approx. 20 ng/μL. The final volume was 45 μL and 1 μL aliquots of the derivatized samples were injected in splitless mode on a HP5-MS 30 m x 0.25 mm x 0.25 μm capillary column (Agilent Technologies, Palo Alto, USA) using a temperature gradient from 70°C to 320°C at a rate of 5°C/min. GC-MS analysis was performed using an Agilent 6890 gas chromatograph coupled to an Agilent 5973 mass selective detector (Agilent Technologies). Detection was carried out using MS detection in electron impact mode and full scan monitoring mode (m/z 15-800). The electron impact for the generation of ions was 70 eV.

Sample work-up was carried out in duplicate for all samples. A pooled human CSF sample was analyzed in sextuplicate to determine the analytical variation in the analysis of metabolites by GC-MS. Data-pre-processing was carried out by composing target lists of peaks detected in the samples based on retention time and mass spectra and these peaks were integrated for all samples. All peak areas were subsequently normalized using internal standards. The resulting target lists were used for further statistical analysis. Identities were assigned based on the presence of identical mass spectra in an in-house database.

LC-MS/MS analysis

To 10 μL of human CSF sample, 10 μL of an internal standard solution containing $^{13}C^{15}N$-amino acids was added followed by addition of 100 μL of MeOH. The mixture was vortexed for 10 s and centrifuged at 10.000 rpm for 10 min at 10 °C. The supernatant was dried under N_2. The residues were dissolved in 80 μL borate buffer (pH 8.5) and after 10 s vortexing 20 μL of AQC reagent (Waters, Etten-Leur, The Netherlands) was added and the mixture was vortexed immediately. The samples were heated 10 min at 55°C. After cooling down, a 1μL sample of the reaction mixture was injected into the UPLC-MS/MS system.

An ACQUITY UPLC™ system with autosampler (Waters) was coupled online with a Quattro Premier XE Tandem quadrupole mass spectrometer (Waters) and was used in positive-ion electrospray mode. The instrument was operated under Masslynx data acquisition software (version 4.1; Waters). The samples were analyzed by UPLC-MS/MS using a AccQ-TagTM Ultra 100 mm x 2.1 mm (1.7μm particle size) column (Waters). A binary gradient system of water – eluent A (10:1, v/v) (AccQ Tag, Waters) and 100% eluent B (AccQ Tag, Waters), was used. Elution of the analytes was achieved by ramping the percentage of eluent B from 0.1 to 90.0 in approx. 9.5 minutes using a combination of both linear and convex profiles. The flow-rate was 0.7 mL/min. the column temperature was maintained at 60°C and the temperature of the autosampler tray was set to 10°C. After each injection the injection needle was washed with 200 μL strong wash solvent (95% ACN), and 600μL weak wash solvent (5% ACN).

The Quattro Premier XE was used in the positive-ion electrospray mode and all analytes were monitored in Selective Reaction Monitoring (SRM) using nominal mass resolution (FWHM 0.7 amu). Next to the derivatisation reagent all amino acids were selectively monitored via the transition from the protonated molecule of the AccQ-Tag derivative to the common fragment at m/z 171. Collision energy and collision gas (Ar) pressure were 22eV and 2.5 mbar, respectively. The complete chromatogram was divided into 6 time windows, restricting the number of SRM transitions to follow and allowing quantitative information to be gathered in each segment. Acquired data was evaluated using Quantlynx (Waters). All samples were analyzed in duplicate.

Data-pre-processing was carried out by calculating the concentration of 18 amino acids in all samples by peak integration, followed by normalization using relevant internal standards and quantification using external calibration curves. The analytical variation was determined from the duplicate analysis of the samples using weighted regression [27, 28]

NMR analysis
CSF samples (280 µL) were centrifuged (2000g, 15 minutes) using a filter with a cut-off of 10 kDa (Centrisart I 13239-E) to remove proteins. Next, 25 µL of 8.8 mM TSP-d_4 stock solution in D_2O was added to 250 µL filtrated CSF to a final concentration of 0.8 mM TSP. The pH of the filtrated CSF was adjusted to around 7 (7.0 – 7.1) by adding phosphate buffer (9.7 µL 1M, to a final concentration of 35 mM). The final CSF NMR sample (284.7 µL) was then transferred to a Shigemi microcell NMR tube for NMR measurements (called non-diluted CSF samples). As a duplicate, for establishing the analytical variation, 100 µL of CSF of the individuals was diluted into 180 µL $D_2$0 and subsequently worked up as described above (called further on diluted CSF samples).

The 1D ^{1}H NMR spectra of diluted and non-diluted CSF samples were acquired on an 800 MHz Inova (Varian, city, country) system equipped with either a 5 mm triple-resonance, XYZ-gradient HCN room-temperature probe or a 5 mm triple-resonance, Z-gradient HCN cold-probe, respectively. Suppression of water was achieved by using WATERGATE (delay: 85 µs) [29]. For each 1D ^{1}H NMR spectrum 512 scans of 18K data points were accumulated with a spectral width of 9000 Hz. The acquisition time for each scan was 2 s. Between scans an 8 s relaxation delay was employed. Prior to spectral analysis, all acquired Free Induction Decays (FIDs) were zero-filled to 64K data points, multiplied with a 0.3 Hz line broadening function, Fourier transformed and manually phase - and baseline corrected by using ACD/SpecManager software. Spectra were subsequently transformed to the Chenomx NMR Suite Professional software package version 5.1 for further analysis [30]. Metabolite identification and quantification were done by using the 800 MHz library of metabolite NMR spectra from the Chenomx NMR Suite 5.1 (pH 6-8). The metabolite spectra in the library are predicted based on a database of pure compound spectra acquired using particular pulse sequence and acquisition parameters, e.g. the tn-noesy-presaturation pulse sequence with 4s acquisition time and 1s of recycle delay. The Chenomx NMR Suite software fits the spectral signatures (singlets, doublets, triplets etc), i.e. the peak shapes, of a compound from an internal database of reference spectra to the experimental NMR spectrum. The resonance assignments derived from the Chenomx NMR Suite software were further checked against literature spectra. For quantification, Chenomx NMR Suite 5.1 uses the concentration of the known reference signal as calibration (in this case TSP-d_4).

The analytical variation on the individual metabolite concentrations was determined from the NMR analysis of the dilute and non-dilute CSF samples, completely independently, and the quintuplicate measurement of the diluted CSF sample of one individual.

Biological variation in an experimental setting

To test the results on the variation of metabolite abundances found in the nine individual CSF samples, an identical experiment was performed using a larger set of samples analyzed by GC-MS. A total of 42 human CSF samples were used, taken from patients with multiple sclerosis and other (inflammatory) neurological diseases.

RESULTS

Proteomics

None of the CSF samples was contaminated with plasma, as according to the criteria set hemoglobin and apolipoprotein B100 were not identified in any of the samples. All sequenced peptides and identified proteins are listed in the supplementary material (including the number of unique peptides per protein and the sequence coverage for all proteins identified with two or more peptides).

Using MALDI-FT-ICR mass spectrometry we analysed the height of albumin peptide peaks and their correlation to albumin concentration levels in CSF as measured by routine clinical chemistry diagnostics. The sum of the height of 14 omnipresent albumin peaks showed positive correlation to the values measured by

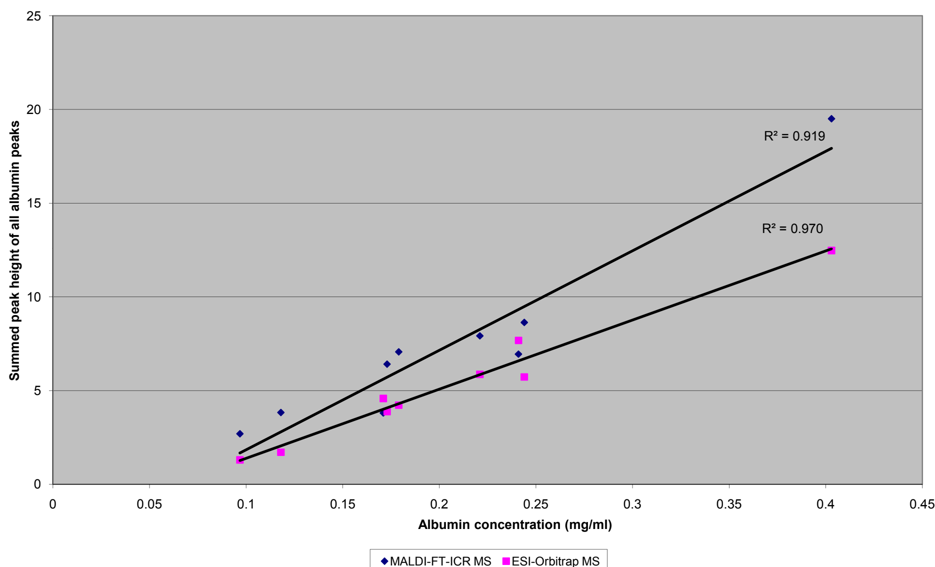


Figure 1: Correlation between the measured albumin concentration by clinical chemistry diagnostics and the sum of the height of 14 omnipresent albumin peaks as measured by MALDI-FT-ICR (R^2 = 0.919) and by ESI-Orbitrap (R^2 = 0.971).

clinical chemistry (R^2 = 0.919). These values (median: 0.219 g/L, range 0.097-0.403 g/L) clearly show a large variation between individuals, which was also apparent from the differences in height of the peaks measured by MALDI-FT-ICR (Figure1). The area under the curve of all peptides identified to be part of albumin in the ESI-Orbitrap experiments was also plotted against the albumin concentration, showing good correlation (R^2 = 0.971). Relative standard deviations (RSD) were comparable for all three methods (43.7% for clinical chemistry, 66.7% for MALDI-FT-ICR and 39.1% for ESI-Orbitrap).

A total of 130 proteins, all identified by multiple peptides and present in all 9 normal CSF samples, was analysed in the nine individual CSF samples by ESI-Orbitrap to assess the variance in protein abundances in CSF, based on the averages of peak heights of all the peptides of a single protein. The RSD ranged from 18% to 148% (median: 43%) in peak height. The far greater part of the examined proteins (123 of 130, i.e. 94.6%) showed lower than 100% RSD in average peptide peak height per protein between the nine individual CSF samples. These results were subsequently tested in a larger, experimental sample set (36 samples), in which a similar profile for the variation in protein abundance, based on the averages of the area-under-the-curve of the peptides in the ESI-Orbitrap, was observed (Figure 2). In the original sample set all proteins were observed in all 9 normal control

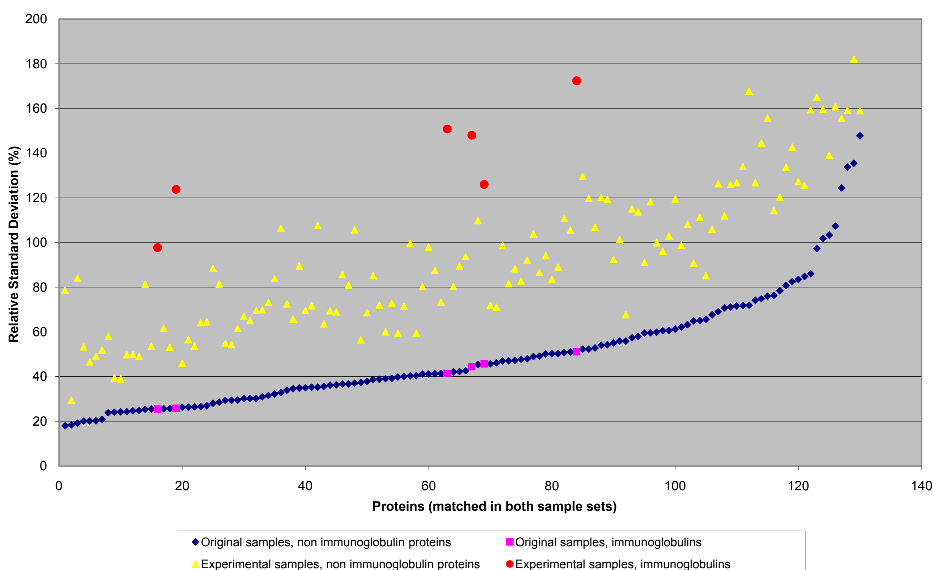


Figure 2: The proteins sorted by the variation in the original samples paired to the variation in the experimental samples. A trend is clearly visible, but due to the nature of the experimental samples (multiple sclerosis and headaches), the immunoglobulins do not correspond with the overall trend, which is to be expected considering the well-known inflammatory component of multiple sclerosis.

samples, but in the experimental sample set 11 proteins were not observed in all 36 samples, providing a first indication of a greater variance in the these samples. In the experimental sample set, the RSD in peptide abundance per protein ranged from 30% to 182% (median: 91%). The greater variation in the experimental sample set is at least partially due to the sample choice for this set of samples. As referenced earlier, in multiple sclerosis it is known that immunoglobulins are elevated and since this sample set contained both multiple sclerosis CSF samples and samples from patients with headaches it is not surprising to note that many of the proteins with the highest RSD between individuals are all immunoglobulin types and proteins related to inflammatory response, which were indeed elevated in the multiple sclerosis samples. In essence, in the normal controls we observed the biological variation between the individuals, whereas in the experimental sample set both the biological variation as well as the disease-related variation was observed.

Although both sample sets are distinct, a number of similarities can be observed. In both sample sets, there is a clear division that can be seen between proteins whose abundances vary highly among individuals and proteins that show a much more limited variation between individuals. Among the proteins that showed limited variation between individual CSF samples were serotransferrin (25% RSD in the original sample set and 50% RSD in the experimental sample set), fibulin-1 (26% and 64%, respectively), and phospholipid transfer protein (27% and 60%, respectively). Proteins with high variation between individuals in both sample sets included cadherin-13 (82% RSD in the original sample set and 143% RSD in the experimental sample set), contactin-2 (124% and 156%, respectively), and haptoglobin (135% and 182%, respectively). The full list of variations between the individuals for the 130 proteins can be found in the Supplementary Material.

Metabolomics

Three different analytical methods were applied to analyse the individual as well as the pooled CSF samples. The methods included untargeted GC-MS and NMR methods and a targeted LC-MS/MS method specifically for amino acids.

Analysis of CSF with GC-MS resulted in a list of 108 metabolites of which 89 could be identified (Table 2). The unknown metabolites covered both metabolites that were observed in other biofluids, i.e. plasma and/or urine, as well as metabolites that seemed to be specific for CSF. As can be seen from Table 2, the metabolites detected by GC-MS cover many different compound classes, i.e. amino acids, organic acids, nucleosides, fatty acids, mono- and disaccharides. Of the 89 identified metabolites, some were only present in trace amounts and were therefore not used for further analysis. Interestingly, all identified metabolites were observed in

Table 2. Metabolites detected by GC-MS, NMR and LC-MS/MS and their biological variation.

Metabolite	GC-MS RSD (%) n=9	NMR RSD (%) n=5	LC-MS/MS RSD (%) n=8
1,5-Anhydro-D-Glucitol	56		
1-Methylhistidine		50	
1-Monopalmitoylglycerol	61		
1-Monostearoylglycerol	63		
2,3,4-Trihydroxybutyric acid	27		
2,4-Dihydroxybutyric acid	29		
2-Aminobutyric acid	40	17	
2-Hydroxybutyric acid	38	16	
2-Hydroxyisovaleric acid	33	43	
2-Hydroxypropenoic acid	47		
2-Oxo-butyric acid	33		
3,4-Dihydroxybutyric acid	25		
3-Hydroxybutyric acid	46	50	
3-Hydroxyhexanoic acid	39		
3-Hydroxyisovaleric acid	31	42	
3-Hydroxypropanoic acid	28		
3-Methylhistidine		25	
5,6-Dihydrouracil	28		
Acetic acid		33	
Acetoacetic acid	59	28	
Aconitic acid		46	
Alanine	41	19	40
Aminobutyric acid isomer	48		
Aminomalonic acid	43		
Arabinose	31		
Arabitol	40		
Arginine		9	28
Ascorbic acid	46		
Asparagine	39		34
Aspartic acid	37		55
Benzoic acid	22		
C14:0 Fatty acid	43		
C16:0 Fatty acid	25		
C18:0 Fatty acid	21		
Cholesterol	29		
Choline		22	
Citric acid	44	19	
Creatine		13	
Creatinine	30	12	
Cysteine	30		
Dimethylamine		15	
Erythronic acid	32		
Formic acid		8	
Fructose	35	29	
Fucose	30		
Galactitol		13	

Table 2. (continued)

Metabolite	GC-MS RSD (%) n=9	NMR RSD (%) n=5	LC-MS/MS RSD (%) n=8
Gluconic acid	38		
Glucose	15	16	
Glutamic acid	32		74
Glutamine	28	20	31
Glyceric acid	25		
Glycerol	29		
Glycine	22	34	
Glycolic acid	20		
Histidine		23	32
Hydroxyproline	59		
Hypoxanthine	43		
Inositol	85		
Iso-leucine	43	21	34
Lactic acid	16	12	
Leucine	30	26	29
Lysine	26	37	30
Mannitol	38		
Mannose	31		
Meso-erythrytol	31		
Methanol		27	
Methionine	45	37	46
Myo-inositol	37	34	
Myo-inositol-1,2-cyclicphosphate	46		
	56		
N-Acetylaminomalonic acid			
Ornithine	29		
Phenylalanine	27	20	41
Phosphoric acid	41		
Proline	44		52
Pseudo uridine	37		
Pyruvic acid	34	16	
Ribitol	32		
Ribonic acid or isomer	29		
Ribose	30		
Serine	31		30
sn-Glycerol-3-Phosphate	35		
Succinic acid		23	
Trimethylamine-N-oxide		25	
Threonine	23	17	32
Tryptophan	53		44
Tyrosine	41	22	42
Urea	53	53	
Uric acid	44		
Uridine	42		
Valine	36	18	38
Xanthine		35	

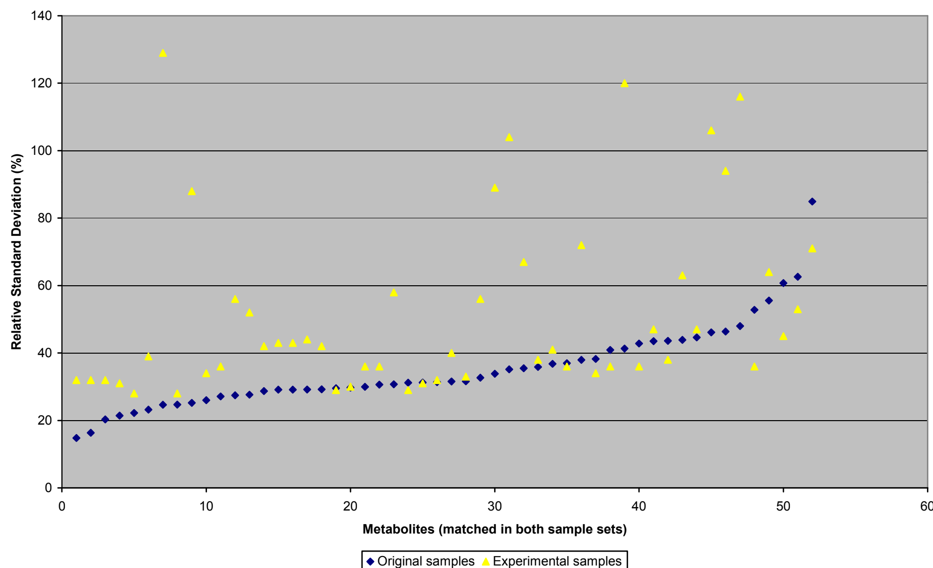


Figure 3: Metabolites (detected by GC-MS in both the original as well as the experimental sample set) sorted by variation in the original samples paired to the variation in the experimental samples. A trend is clearly visible, but due to the nature of the experimental samples (multiple sclerosis and other neurological disorders), a number of metabolites do not correspond with the overall trend and show significantly higher biological variation.

all samples. The analytical variation for each metabolite was determined from the repeated (n=6) analysis of the pooled CSF sample. Results show that the analytical variation (< 20%) was less than the biological variation for all metabolites (15 to 85%) (Table 2). The biological variation for the metabolites observed with GC-MS shows a normal distribution, as can be deduced from Figure 3. The concentration or relative peak area for each metabolite in the pooled human CSF sample is given in Supplemental Material. As expected, the average concentrations and relative peak areas found for the 9 individual CSF samples were very similar to that of the pooled human CSF sample.

For the experimental sample set of 42 human CSF samples from patients having neurological diseases a similar profile for the variation in metabolite level could be observed as for the nine CSF samples from neurologically normal individuals (Figure 3). Note that the number of metabolites plotted in Figure 3 is less than 76 because for the experimental sample set only 40 μL could be used due to limited available sample volume and thus only metabolites were used that could be integrated accurately in both sample sets. A number of metabolites show a significantly higher RSD for the experimental samples, which is most probably due to the heterogeneity of the experimental CSF samples, as discussed in the proteomics section. Metabolites that show significantly higher RSDs for the

experimental samples are 3,4-dihydroxybutanoic acid, fructose, phosphoric acid, 3-hydroxybutanoic acid and 3-aminobutyric acid. However there is no clear relation between these metabolites and the neurological disease in the experimental sample set.

For NMR five individual CSF samples were analysed due to limited available sample volumes. Analysis of the CSF samples by NMR resulted in a list of 51 metabolites of which 42 could be quantified (Table 2). All metabolites observed with NMR were detected in all samples. The biological variation ranged from 8 to 53% while the analytical error was between 3 and 9% for all metabolites (Table 2). The concentrations found for the pooled CSF sample were very similar to that of the averages of the individual samples.

Of the 51 metabolites detected by NMR in this study, 35 were also detected by GC-MS. Some of the more volatile metabolites, like acetone and methanol, can only be analyzed by NMR, showing that despite the overlap, NMR and GC-MS are complementary techniques. Furthermore, NMR can detect a number of metabolites that are difficult to analyze by GC-MS, because they cannot be derivatized, like choline, or they can give unstable derivatives, like arginine. On the other hand, a range of metabolites was only observed by GC-MS and not by NMR. In most cases these metabolites either have no proton signal, e.g. uric acid and phosphoric acid, or the concentration is below the detection limit of NMR, e.g. proline and tryptophan.

The absolute concentrations of amino acids in eight individual CSF samples were determined by a targeted LC-MS/MS. One of the individual CSF samples was omitted due to technical failure. The analytical error is less than the biological variation for all amino acids. The biological variation ranges from 28 to 74% while the analytical error is <20% for all metabolites, except aspartic acid (35%) (Table 2). Despite the differences between samples, all amino acids were present in every individual sample. Again, it can be seen that the concentrations found for the pooled CSF sample were very similar to that of the averages of the individual samples.

Most of the amino acids analyzed by LC-MS/MS were also detected either by GC-MS or NMR. However, one of the advantages of the targeted LC-MS/MS method is the low sample volume required for analysis, i.e. 10 µl vs. 60-100 µl for NMR and GC-MS. Comparison of the RSD of metabolites that could be analyzed with more than one of the analytical methods shows that on average the biological variation of a metabolite is similar for different methods. Deviations occur mainly for low abundant metabolites and metabolites that show low sensitivity for certain methods.

DISCUSSION

In this study, we investigated metabolite and protein identities, and their abundances and inter-individual variations in abundance in CSF by analyzing a unique and well-defined set of CSF samples and a corresponding pooled CSF sample. Here we have strictly defined criteria to exclude blood contaminated CSF. These criteria warrant that at a certain threshold no contamination is observed, however contamination not exceeding this threshold can still exist and cannot be ruled out.

Combination of three different analytical techniques for metabolites used in this study resulted in a list of about 105 identified metabolites in CSF, which is about a third of the metabolites in CSF present in the human metabolome database [30]. It is expected that many of the metabolites that are not detected by NMR and GC-MS are low abundant metabolites, i.e. neurotransmitters, steroids, eicosanoids, for which more specific, targeted methods are required [31-33]. However, these methods often require significant amounts of CSF and should therefore only be used in metabolomics studies when there is evidence that these metabolites are of importance and/or when enough sample volume is available. Furthermore, some metabolites are (almost) absent in normal controls and are only detectable in diseased persons [8].

All endogenous metabolites detected with the three analytical methods in this study were observed in all individual CSF samples. This implies that the qualitative metabolite composition of CSF in normal controls is relatively similar between individuals. This is generally also observed for plasma of healthy persons in contrary to urine, which is actually probably more influenced by dietary intake.

With NMR and LC-MS/MS it was possible to determine the absolute concentration of metabolites. This in contrast to GC-MS for which metabolites can only be quantified when either internal standards or calibration curves for each metabolite are used, which is practically not feasible and therefore this method, like many other non-targeted methods, is used to measure relative differences in metabolite concentrations between groups or individuals. The concentrations found for amino acids with LC-MS/MS and NMR can be compared with values found in normal control CSF samples in literature [13, 34-36].

Although the study of Hu *et al.* [16] mainly focussed on the variation of specific protein abundances within individuals, they concluded that inter-individual variation is far more extensive than intra-individual variation. Yet in that study two different stages of Alzheimer's disease were included, which could potentially influence the levels of protein variation between individuals. Here we examined nine well-defined CSF samples taken from patients without neurological disorders, and also found profound differences in protein abundances between individuals. Characterization

of variation of CSF levels of amyloid beta [37] and apolipoprotein E [38] in patients with Alzheimer's disease have been published, but this is the first attempt to characterize a large number of proteins in CSF of patients without neurological afflictions. Some proteins, such as serotransferrin and fibulin-1 appear to be more constant than others with regards to abundance levels in CSF, as these showed only limited variation in both the sample set of nine non-neurological individuals as well as in the experimental samples. Other proteins, such as contactin-2 and cadherin-13, showed large variations in abundance levels in both data sets, while proteins related to inflammatory response showed the largest variation in the experimental sample set (Figure 2). This is, in all likelihood, due to the well-known neuroinflammatory component of multiple sclerosis [39-41], because the abundance of neuroinflammatory proteins was far higher in the multiple sclerosis samples.

Additionally it must be noted that the proteins with high inter-individual variation were not only low abundant proteins, but also high abundant proteins such as haptoglobin, indicating that these high variations were not caused by measuring at the limit of the detection capabilities of the machines.

Although all metabolites detected were present in all individual samples the concentration of metabolites differed strongly between individuals. For all metabolites the analytical variation was significantly less than the biological variation. The biological variation in this study is about 30% for 60% for the majority of the metabolites which is significantly lower than was observed by Wishart *et al.* [13], but very similar to the variation reported in the recent study of Crews *et al.* [42]. The main difference between the two studies is the type of CSF sample that was used, i.e. persons without neurological disorders vs. patients screened for meningitis. Therefore, it can be concluded that the biological variation for normal controls is, as expected, less than for neurologically diseased individuals.

Lactic acid and glucose, two high abundant metabolites that can be detected by both NMR and GC-MS, show relatively low biological variation, i.e. ~15%, as well as other metabolites detected by NMR, i.e. formic acid, creatine and creatinine (Table 2). On the contrary, some high abundant metabolites, like fructose, urea and myo-inositol, already show much higher biological variations ranging from 30-50% with both GC-MS and NMR (Table 2). Somehow these high-abundant metabolites vary much more between individuals. Other metabolites that show high biological variations are aconitic acid, 3-hydroxybutyric acid and ascorbic acid (Table 2). There is yet no clear biological reason why some metabolites in this study show much higher biological variation than others. The biological variation of the experimental sample set of 42 human CSF samples showed a similar trend as the nine CSF samples, although for a number of metabolites the biological variation was significantly higher (Figure 3). The latter can of course be attributed to the

different diseases of the subjects used that lead to differences in metabolite levels. Metabolites that showed a significant higher biological variation in the experimental samples could not be directly related to neurological disorders. Further analysis of the data of the experimental samples is necessary in order to find relations between metabolites and the different types of neurological disorders present in the samples set, including the different stages of MScl.

More interestingly, for a significant number of metabolites, the biological variation in diseased subjects are similar to that of normal controls, indicating that part of the CSF metabolome is more influenced by person to person differences and that the contribution of diseases is only minor.

The biological variation of metabolites that were detected by more than one analytical method, showed in general good agreement. For example, the difference in biological variation of amino acids observed by GC-MS and LC-MS/MS was less than 10% for most metabolites. Exceptions were aspartic acid and glutamic acid which were present at low concentrations and which showed relatively high analytical variation with LC-MS/MS. A similar comparison between metabolites detected with both NMR and GC-MS, i.e. lactic acid, fructose, glucose, myo-inositol, pyruvic acid, urea, showed the same good agreement.

The work discussed above showed that metabolomics, i.e. non-targeted analysis of as many metabolites as possible, of CSF is possible with a combination of analytical techniques currently available. Depending on the amount of CSF available and existing knowledge with respect to the biological question that has to be answered, a combination of non-targeted and targeted analytical methods is preferred to cover different classes of metabolites ranging from high to low abundant metabolites. The metabolites that were detected in CSF seemed to be quite similar between normal controls although the concentration of metabolites can differ between individuals up to 60% depending on the specific metabolite.

CONCLUDING REMARKS

From the previous discussion we conclude that for most proteins the biological variation in the nine individual normal control CSF samples, i.e. patients without any significant neurological disorders, appears to be limited, e.g. serotransferrin with RSD 25%, which includes a technical variation of approximately 20%. The majority of the identified proteins show lower than 60% RSD. However for 28% of the identified proteins the RSD is above 60% and for a limited number of proteins (5% of total) the inter-individual variation is extensive (RSD > 100%). The results of extensive inter-individual variation for 5% of the identified proteins is not limited

Table 3. Proteins with high and low biological variation in normal control CSF samples.

Low variation between individuals			High variation between individuals		
Accession number, protein	Biological variation (%)	Technical variation (%)	Accession number, protein	Biological variation (%)	Technical variation (%)
P01011, Alpha-1-antichymotrypsin	26	17	Q9BYH1, Seizure 6-like protein	102	21
P07339, Cathepsin D	26	17	Q8TCZ2, Voltage-dependent calcium channel subunit alpha-2/delta-1	103	19
P23142, Fibulin-1	27	16	P54764, Ephrin type-A receptor 4	107	19
P02774, Vitamin D-binding protein	29	21	Q02246, Contactin-2	124	19
P17900, Ganglioside GM2 activator	29	23	P00738, Haptoglobin	135	28
P02749, Beta-2-glycoprotein	30	18	Q86YZ3, Hornerin	148	28

to low abundant proteins but also several high abundant proteins shows extensive biological variation, e.g. haptoglobin with RSD 135% (Table 3).

Metabolomics analysis on the same CSF samples showed that the biological variation for most CSF metabolites is limited especially compared to the proteomics results. No metabolites were observed with a biological RSD > 80%. However, within the group of metabolites that could be quantified with the different analytical methods, substantial differences in RSDs could be observed between individual metabolites. For example, glucose, a high abundant metabolite, showed a RSD of only 15% while for fructose, another high abundant carbohydrate, a RSD of 35% was observed.

These results show that for CSF biomarker discovery research, it is essential to have an understanding of the biological variation between normal controls, because observation of differential abundance between controls and diseased individuals must necessarily be weighed against known inter-individual variations in normal controls. Proteins and metabolites showing high RSD in healthy CSF ought to be assessed with caution as candidate biomarker, since a large part of the observed difference will not be due to the disease under investigation, but to the natural biological variation between individuals.

ACKNOWLEDGEMENTS

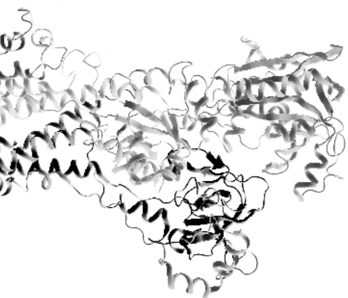
This study was performed within the framework of Top Institute Pharma project number D4-102.

REFERENCES

1. Frankfort, S.V., et al., *Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia: a review of recent literature.* Curr Clin Pharmacol, 2008. **3**: p. 123-31.
2. Helbok, R., et al., *Chronic meningitis.* J Neurol, 2009. **256**: p. 168-75.
3. Lewczuk, P., et al., *Neurochemical dementia diagnostics: assays in CSF and blood.* Eur Arch Psychiatry Clin Neurosci, 2008. **258 Suppl 5**: p. 44-9.
4. Romeo, M.J., et al., *CSF proteome: a protein repository for potential biomarker identification.* Expert Rev Proteomics, 2005. **2**: p. 57-70.
5. Johnston, I., et al., *Disorders of CSF hydrodynamics.* Childs Nerv Syst, 2000. **16**: p. 776-99.
6. Taniguchi, M., et al., *Sugar chains of cerebrospinal fluid transferrin as a new biological marker of Alzheimer's disease.* Dement Geriatr Cogn Disord, 2008. **26**: p. 117-22.
7. Baran, R., et al., *Mass spectrometry based metabolomics and enzymatic assays for functional genomics.* Curr Opin Microbiol, 2009.
8. Moolenaar, S., von Engelke, U., Hoenderop, S., Morava, E., van der Graaf, M., Wevers, R., *Handbook of ^{1}H-NMR spectroscopy in inborn errors of metabolism.* 2002: Heilbronn: SPS verlagsgesellschaft.
9. Lindon, J.C., Nicholson, J. K., Everett, J. R., *NMR spectroscopy of biofluids*, in *Annual reports on NMR spectroscopy*, G.A. Webb, Editor. 1999, Academic Press: London. p. 1-88.
10. Lutz, N.W., et al., *Inflammatory multiple-sclerosis plaques generate characteristic metabolic profiles in cerebrospinal fluid.* PLoS One, 2007. **2**: p. e595.
11. Kawashima, H., et al., *Primary biomarkers in cerebral spinal fluid obtained from patients with influenza-associated encephalopathy analyzed by metabolomics.* Int J Neurosci, 2006. **116**: p. 927-36.
12. Myint, K.T., et al., *Quantitative profiling of polar cationic metabolites in human cerebrospinal fluid by reversed-phase nanoliquid chromatography/mass spectrometry.* Anal Chem, 2009. **81**: p. 1121-9.
13. Wishart, D.S., et al., *The human cerebrospinal fluid metabolome.* J Chromatogr B Analyt Technol Biomed Life Sci, 2008. **871**: p. 164-73.
14. Pan, S., et al., *A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry.* Proteomics, 2007. **7**: p. 469-73.
15. Pan, S., et al., *Application of targeted quantitative proteomics analysis in human cerebrospinal fluid using a liquid chromatography matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometer (LC MALDI TOF/TOF) platform.* J Proteome Res, 2008. **7**: p. 720-30.
16. Hu, Y., et al., *Comparative proteomic analysis of intra- and interindividual variation in human cerebrospinal fluid.* Mol Cell Proteomics, 2005. **4**: p. 2000-9.
17. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.* Mol Cell Proteomics, 2005. **4**: p. 2010-21.
18. Stoop, M.P., et al., *Quantitative matrix-assisted laser desorption ionization-fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple-sclerosis-related proteins.* J Proteome Res, 2009. **8**: p. 1404-14.
19. Mize, T.H., et al., *Broad-band ion accumulation with an internal source MALDI-FTICR-MS.* Anal Chem, 2000. **72**: p. 5886-91.

20. Moyer, S.C., et al., *Attomole peptide analysis by high-pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry*. Anal Chem, 2003. **75**: p. 6449-54.
21. O'Connor, P.B., et al., *Application of multishot acquisition in Fourier transform mass spectrometry*. Anal Chem, 2000. **72**: p. 5125-30.
22. van Kampen, J.J., et al., *Quantitative analysis of HIV-1 protease inhibitors in cell lysates using MALDI-FTICR mass spectrometry*. Anal Chem, 2008. **80**: p. 3751-6.
23. Frequin, S.T., et al., *CSF myelin basic protein, IgG and IgM levels in 101 MS patients before and after treatment with high-dose intravenous methylprednisolone*. Acta Neurol Scand, 1992. **86**: p. 291-7.
24. Presslauer, S., et al., *Elevated levels of kappa free light chains in CSF support the diagnosis of multiple sclerosis*. J Neurol, 2008. **255**: p. 1508-14.
25. Williams, K.C., et al., *Immunology of multiple sclerosis*. Clin Neurosci, 1994. **2**: p. 229-45.
26. Koek, M.M., et al., *Microbial metabolomics with gas chromatography/mass spectrometry*. Anal Chem, 2006. **78**: p. 1272-81.
27. Massart, L.M., Vandenginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics, Part A*. 1997, Amsterdam: Elsevier.
28. [cited; Available from: http://cran.r-project.org/web/packages/chemCal/index.html.
29. Sweatman, B.C., et al., *600 MHz 1H-NMR spectroscopy of human cerebrospinal fluid: effects of sample manipulation and assignment of resonances*. J Pharm Biomed Anal, 1993. **11**: p. 651-64.
30. Wishart, D.S., et al., *HMDB: the Human Metabolome Database*. Nucleic Acids Res, 2007. **35**: p. D521-6.
31. Kim, Y.S., et al., *Profiling neurosteroids in cerebrospinal fluids and plasma by gas chromatography/electron capture negative chemical ionization mass spectrometry*. Anal Biochem, 2000. **277**: p. 187-95.
32. Obata, T., et al., *Simultaneous assay of prostaglandins and thromboxane in the cerebrospinal fluid by gas chromatography-mass spectrometry-selected ion monitoring*. J Chromatogr B Biomed Sci Appl, 1999. **731**: p. 73-81.
33. Rodriguez, S., et al., *Changes in the cerebrospinal-fluid monoamines in rats with an immunoneutralization of the subcommissural organ-Reissner's fiber complex by maternal delivery of antibodies*. Exp Brain Res, 1999. **128**: p. 278-90.
34. Fonteh, A.N., et al., *Quantification of free amino acids and dipeptides using isotope dilution liquid chromatography and electrospray ionization tandem mass spectrometry*. Amino Acids, 2007. **32**: p. 203-12.
35. Molina, J.A., et al., *Cerebrospinal fluid levels of non-neurotransmitter amino acids in patients with Alzheimer's disease*. J Neural Transm, 1998. **105**: p. 279-86.
36. Qureshi, G.A., et al., *Quantitation of free amino acids in biological samples by high-performance liquid chromatography. Application of the method in evaluating amino acid levels in cerebrospinal fluid and plasma of patients with multiple sclerosis*. J Chromatogr, 1988. **459**: p. 237-44.
37. Bateman, R.J., et al., *Fluctuations of CSF amyloid-beta levels: implications for a diagnostic and therapeutic biomarker*. Neurology, 2007. **68**: p. 666-9.
38. Lefranc, D., et al., *Relevance of the quantification of apolipoprotein E in the cerebrospinal fluid in Alzheimer's disease*. Neurosci Lett, 1996. **212**: p. 91-4.
39. Compston, A., et al., *Multiple sclerosis*. Lancet, 2002. **359**: p. 1221-31.

40. Frohman, E.M., et al., *Immunologic mechanisms of multiple sclerosis.* Neuroimaging Clin N Am, 2008. **18**: p. 577-88, ix.
41. Uccelli, A., et al., *Biological markers of the inflammatory phase of multiple sclerosis.* Neurol Sci, 2003. **24 Suppl 5**: p. S271-4.
42. Crews, B., et al., *Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data.* Anal Chem, 2009. **81**: p. 8538-44.



Chapter 3

The rate of false positive sequence matches of peptides profiled by MALDI MS and identified by MS/MS

Marcel P. Stoop, Robert-Jan A. N. Lamers, Peter C. Burgers, Peter A. E. Sillevis Smitt, Rogier Q. Hintzen, and Theo M. Luider

Journal of Proteome Research (2008), 7: 4841-4847.

ABSTRACT

In most MALDI peptide profiling cases, sequencing is required to identify peptides of interest, preferentially by using different mass spectrometry techniques. Using identical samples, we determined the number of false positive matches in sequence of peptide identification using different mass spectrometers. This paper demonstrates that the reliability of the identification phase greatly benefits from concerted MS-technologies and determines the influence of mass accuracy, signal-to-noise and statistical score on peptide identification.

INTRODUCTION

Biomarker discovery is an important topic in current scientific research. The process of biomarker discovery can be approached in several ways, one of which is peptide profiling in disease-afflicted body fluids or tissues using advanced mass spectrometry techniques [1]. In this peptide profiling strategy, fast mass spectrometry techniques are preferable, since, in most cases, a large number of complex samples needs to be measured to enable essential statistical analysis. Automated MALDI-TOF mass spectrometry is a fast technique, with measurement times ranging from ten to thirty seconds per sample, capable of measuring large numbers of samples of both diseased and control patients. Once all samples have been measured and processed, analysis of the mass spectra may lead to a list of peptide masses that are differentially abundant between the pre-defined groups. To identify the differentially abundant peptides, a variety of tandem mass spectrometry techniques can be employed [2, 3].

Biomarker discovery studies have been performed on a multitude of body fluids and tissue types, where the choice of biological matrix was based on the disease under investigation [4]. Using MALDI-TOF profiling, differentially abundant peptides and proteins have been identified by MS/MS for a number of diseases, such as Alzheimer's disease [5], oral cancer [6], Down syndrome [7] and multiple sclerosis [8]. The resulting MS/MS spectrum of fragments is compared to known peptide fragments in a database and a probability for correct identification is assigned to this particular peptide for a certain database [9]. This process is based on statistics, which means false positive identifications can theoretically occur merely by chance. Incorporation of extra parameters in the probability scoring method has been shown to improve the reliability of peptide identification [10].Combining data of different mass spectrometric techniques could be an intriguing way not only to increase the chance of identification of the differentially abundant peptides but also to enhance the reliability of peptide and protein identifications. For example, adding an accurate mass determination of a parent ion obtained by MALDI-FT-ICR increases the mass accuracy of a MALDI-TOF parent ion. Combining this accurate mass of the parent ion to the MS/MS fragments obtained by for instance MALDI-TOF/TOF or an ion trap measurement increases the reliability of the ensuing peptide search in the chosen database.

MATERIALS AND METHODS

Experimental design

In an effort to assess the false positive match in sequence of identifications in combining peptide profiling and identification we conducted an experiment using different spectrometric techniques for ion production (MALDI and ESI) and for mass measurement (TOF and Orbitrap). Fifteen cerebrospinal fluid (CSF) samples were measured and analyzed using both off-line nano-LC MALDI-TOF/TOF as well as online nano-LC ESI-Orbitrap mass spectrometry. The same sample that was measured by MALDI-TOF/TOF was also measured by MALDI-FT-ICR and so the exact mass can be added to the peptide identified by MALDI-TOF/TOF. After the measurements the identifications obtained by MALDI-TOF/TOF method were compared to data obtained by ESI-Orbitrap; all identifications obtained by MALDI-TOF/TOF that were also observed by ESI-Orbitrap were classified as positive identifications, while all the identifications obtained by MALDI-TOF/TOF and MALDI-FT-ICR that were identified as different sequences by ESI-Orbitrap at the same mass window (2 ppm) were classified as false positive matches in sequence. Other factors, such as signal-to-noise ratio, were taken into account, which resulted in a signal-to-noise dependent false positive match in sequence rate of identifications obtained by MALDI-TOF/TOF peptide profiling, combined with a parent ion mass accuracy obtained by MALDI-FT-ICR.

Sample collection and pre-treatment

The method used for CSF sampling from patients at the Department of Neurology of the Erasmus University Medical Centre has been described previously [8]. Fifteen CSF samples (average protein concentration: 0.399 ± 0.106 mg/ml) were selected for nano-LC fractionation and subsequent mass spectrometry measurements. These samples were randomly taken from our CSF sample base, disregarding all known sample parameters except for total protein concentration, for which the samples were matched. These 15 samples were digested by trypsin according to the protocol we have described previously [8]. After digestion all samples were split into two equal volumes. One volume part was used for immediate, online measurement on the nano-LC-ESI-Orbitrap mass spectrometer. The other part was used for offline measurement using nano-LC-MALDI-TOF/TOF. After immediate spotting on a pre-spotted MALDI target plate (PAC 384 plate, Bruker Daltonics, Germany), the fractionated CSF samples were subsequently measured by MALDI-TOF/TOF.

Chromatography and mass spectrometry

In the offline nano-LC-MALDI-TOF/TOF method, one microliter of the digested CSF was injected into the nano-LC (Ultimate3000 system, Dionex, USA) to separate the digested CSF peptides. The separation of the peptides took place on a monolithic column (Dionex, the Netherlands). This was followed by spotting on a MALDI target and subsequently measured by off-line MALDI-TOF/TOF mass spectrometry. Peptide separation was obtained during a 115 min LC run, using a 80 minute gradient of a continuous increasing concentration of acetonitrile (ACN) using a mixture of Solvents A and B (Solvent A: 95% water, 0.1% trifluoroacetic acid (TFA), 5% ACN (v/v); Solvent B: 20% water, 0.1% TFA, 80% ACN, with solvent B increasing from 0% to 50% in 80 minutes). The eluent fractions were spotted onto MALDI target plates pre-spotted with α-cyano-4-hydroxycinamic acid (HCCA) matrix (PAC plates, Bruker Daltonics, Germany). Measurements of the pre-spotted plates were performed by automated MALDI-TOF/TOF (UltraFlex 3, Bruker Daltonics, Germany), using the Warp-LC software package, version 1.1 (Bruker Daltonics, Germany). Mass spectra for all spots were first generated and subsequently MS/MS spectra were acquired in a data dependent manner, using a precursor isolation window of 15 Da. Only peaks that were suitable for MS/MS by a series of pre-set conditions (signal-to-noise ratio (S/N) > 5, most intense peak in window of 5 Da) were analyzed. Identifications were obtained by database searches of the MS/MS spectra using Mascot search program, (version 2.2, Matrix Science, United Kingdom) and the Swiss-Prot database (UniProtKB/Swiss-Prot Release 54.7), using the standard settings of the Warp-LC software package.

One microliter of the above mentioned digested CSF was also loaded onto a nano-LC (Ultimate3000 system, Dionex, USA) linked to an Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Germany). The data-dependent acquisition mode for the MS/MS identification step occurred in the ion trap. The peptides were separated on a C18 PepMap 100 column (15 cm length, internal diameter 75 micrometer, Dionex, USA) after trapping the samples for 7.5 minutes on a C18 PepMap trapping column (5 cm length, internal diameter 300 micrometer), using a flow of 250 nl/min. A 160-minute chromatography run was performed that included an 80-minute continuous gradient of increasing concentrations of ACN, going from 100% solvent A to 55% solvent B (Solvent A: 2% ACN, 98% water, 0.1% formic acid; Solvent B: 80% ACN, 20% water, 0.1% formic acid). The samples were measured online on an ESI-Orbitrap, using a precursor isolation window of 2 Da for MS/MS. The MS/MS identifications were obtained from IPI-database (version 3.35), using the Bioworks 3.2 software package (Thermo Fisher Scientific, Germany) and its SEQUEST feature. The HUPO criteria, with XC scores of 1.8, 2.5 and 3.2 for single, double and triple charged ions, respectively were used to assess the quality of the MS/MS identifications.

Analysis

Identical samples were compared for peptide masses that were identified by both the MALDI-TOF/TOF and the ESI-Orbitrap method. In an effort to ascertain the positive rate of identifications using different mass spectrometry devices with low parts-per million (ppm) mass accuracies, the peptides identified by both methods were subsequently individually compared. By spotting 2,5-dihydroxy benzoic acid (DHB) matrix solution on the PAC-plate and re-measuring the identical spots of interest by MALDI-FT-ICR, we added the exact masses to the peptides identified in the MALDI-TOF/TOF [11]. The mass accuracy of the masses observed in the MALDI-FT-ICR ought to be in a window less than 2.0 ppm of the calculated value of the peptide identified in the very same spot that was investigated by MALDI-TOF/TOF. Calibration of the MALDI-FT-ICR spectra was done externally by measuring the calibration mix spots on the pre-spotted target plate, adjacent to the spot of the measured sample. We also compared the false positive match in sequence rate of identifications using MALDI-FT-ICR mass accuracy windows of 1.0 ppm and 0.7 ppm, which can be obtained by averaging multiple measurements [12]. For the Orbitrap a standard mass window of less than 2.0 ppm of the calculated value was used.

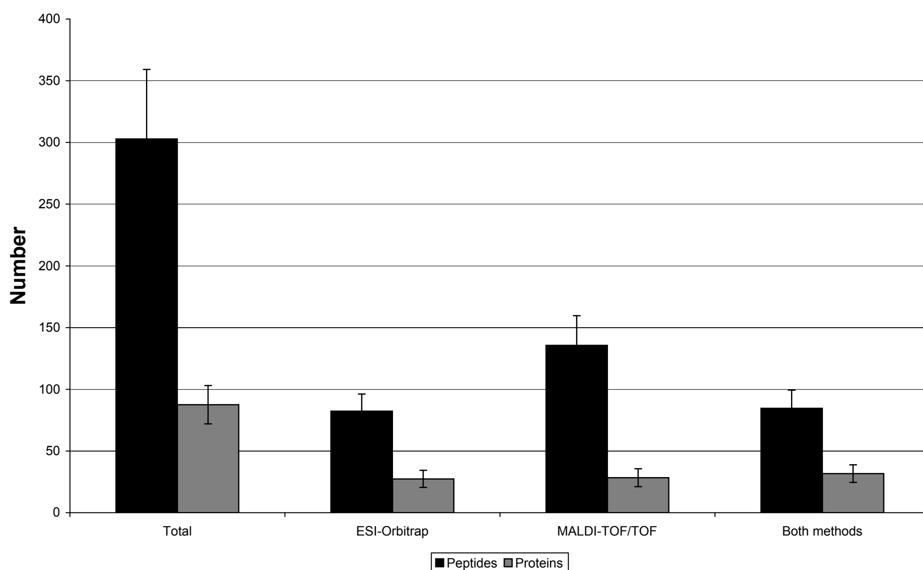


Figure 1. The average number of identified peptides and proteins per sample, using 0.7 ppm mass accuracy in the MALDI-FT-ICR, respectively. The first column denotes the average of all identified peptides and proteins per CSF sample (employing both mass spectrometers), the second column shows the average number of peptides and proteins identified only by ESI-Orbitrap per CSF sample, the third column shows the average number of peptides and proteins identified only by MALDI-TOF/TOF per CSF sample, and the fourth column shows the average number of peptides and proteins identified by both MALDI-TOF/TOF and ESI-Orbitrap per CSF sample.

RESULTS AND DISCUSSION

An average of 303 ± 56 peptides was identified per sample by accumulation of both methods, with an average of 85 ± 14 of these peptides identified by both methods (Figure 1). The average number of proteins identified per sample was 88 ± 16, with an average of 32 ± 7 of these proteins identified by both methods (Figure 1). These numbers set the overlap percentages for peptides and proteins at 28.0% and 36.2%, respectively (Figures 2a and 2b). A total of 249 proteins were identified in this.

Using the MALDI-TOF/TOF method, an average of 221 ± 50 peptides were identified in each of the fifteen CSF samples, which corresponded to an average of 60 ± 13 proteins (Table 1). The ESI-Orbitrap method identified an average of 167 ± 30 peptides using the HUPO criteria, corresponding to an average of 59 ± 10 proteins (Table 1). After re-measuring the PAC plates using the MALDI-FT-ICR, an accurate mass was added to all MALDI-TOF/TOF identifications at a certain spot on the target plate. Restrictions were set to the mass accuracy of the MALDI-FT-ICR (2.0, 1.0 and 0.7 ppm, respectively) as well as the ESI-Orbitrap (2.0 ppm) and subsequently each CSF sample was searched for peptide masses that were identified by both measurement methods. As expected, a more stringent mass accuracy window for the MALDI-FT-ICR resulted in fewer identified peptide peak masses

Table 1. Average peptide and protein identifications of fifteen CSF sample using two different types of mass spectrometry. The MALDI-TOF/TOF column lists the average number of peptides and proteins identified by MALDI-TOF/TOF per sample. The ESI-Orbitrap column lists the same data for the ESI-Orbitrap method. The last column makes an inventory of the average number of proteins and peptides identified by both methods in the CSF samples measured.

	MALDI-TOF/TOF		ESI-Orbitrap		Both Methods	
	Peptides	Proteins	Peptides	Proteins	Peptides	Proteins
Average	221	60	167	59	85	32
Standard deviation	50	13	30	10	14	7

Table 2. Total number of peptide peak masses identified by both MALDI-TOF/TOF and ESI-Orbitrap. Using MALDI-FT-ICR mass accuracy measurements on a machine with a 9.4 T magnet, restrictions of 2.0, 1.0 and 0.7 ppm were set to the accuracy of the parent ion mass identified by MALDI-TOF/TOF.

Total number of overlapping peptides identified by both MALDI-TOF/TOF and ESI-Orbitrap, dependent on mass accuracy window determined by MALDI-FT-ICR	
MALDI-FT-ICR mass accuracy	Number of peptides
2.0	670
1.0	487
0.7	285

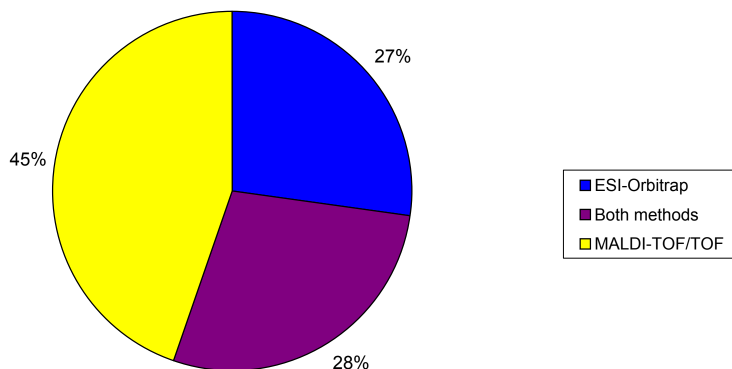


Figure 2a. Pie chart in which the percentages of identified peptides are indicated by the two methods used (MALDI-TOF/TOF or ESI-Orbitrap). A window of 0.7 ppm mass accuracy (MALDI-FT-ICR) was taken into account for these calculations.

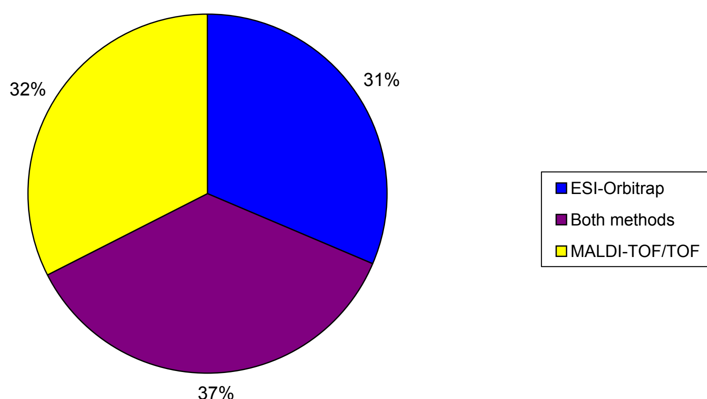


Figure 2b. Pie chart in which the percentages of identified proteins are indicated by the two methods used (MALDI-TOF/TOF or ESI-Orbitrap). A window of 0.7 ppm mass accuracy (MALDI-FT-ICR) was taken into account for these calculations.

that could be matched to ESI-Orbitrap peptide peak masses (Table 2). Peptide peak masses that were identified as different peptides by the two measurement methods were classified as false identifications, while peptides peak masses identified by both methods as the same peptide were classified as correct identifications. The resulting identifications were sorted by S/N ratio (defined as the height of the mass peak above its baseline relative to the standard deviation of the noise) in the MALDI-TOF/TOF spectra. From these data it became apparent that the percentage of false identifications increases as the S/N ratio decreases (Table 3, Figure 3). However, this was only the case with peaks having relatively low S/N ratios.

Although the S/N ratio of the parent peak has an important influence on the result of the MS/MS measurement, it is not the only factor influencing this process. Other factors, such as fragmentation efficiency of a peptide are also critical

Table 3. Total number of identified peptides peak masses found by both MALDI-TOF/TOF and ESI-Orbitrap at various mass accuracy windows and signal-to-noise ratio. The false columns report the number that was not identical when comparing both methods (false positive matches in sequence). The table relates to figure 3 and 4, in which the data is represented graphically and related to identification score, respectively.

S/N	MALDI-FT-ICR mass accuracy					
	2.0 ppm		1.0 ppm		0.7 ppm	
	Total	False	Total	False	Total	False
<10	13	7	8	4	6	2
10-20	19	10	14	6	10	3
20-30	15	6	10	4	8	2
30-40	21	5	15	3	13	1
40-50	6	0	5	0	2	0
50-60	13	0	12	0	8	0
60-70	13	0	10	0	8	0
70-80	9	0	7	0	6	0
80-90	9	1	9	1	4	0
90-100	13	0	9	0	9	0
100-110	11	0	8	0	5	0
110-120	9	0	6	0	6	0
120-130	9	0	5	0	2	0
130-140	17	0	12	0	8	0
140-150	10	0	9	0	2	0
150-160	7	0	5	0	3	0
160-170	7	0	4	0	3	0
170-180	11	0	10	0	6	0
180-190	5	0	4	0	3	0
190-200	13	0	12	0	7	0
>200	440	0	313	0	184	0

for peptide identification. The Mascot score describes the probability of a correct identification (using MALDI-TOF/TOF). If the score is low, then the probability of a false positive match in sequence is higher. However, the false identifications of peptide peak masses we found did not have significantly lower Mascot scores than correct identifications when compared in sub-entities grouped by S/N ratio ($p = 0.29 – 0.97$, Figure 4). When all Mascot scores are compared, regardless of S/N ratio, the p-value of a 2-tailed t-test comparing correct and false identifications is statistically significant ($p = 0.0107$).

Observing these results it becomes increasingly obvious that multiple peptide identification strategies increase the reliability of the identifications, as conflicting data for any single identification can be examined in detail and discarded by objective measures. Two important observations emerge from the present study. First, the number of false positive matches in sequence appears to be low (i.e. 8/303,

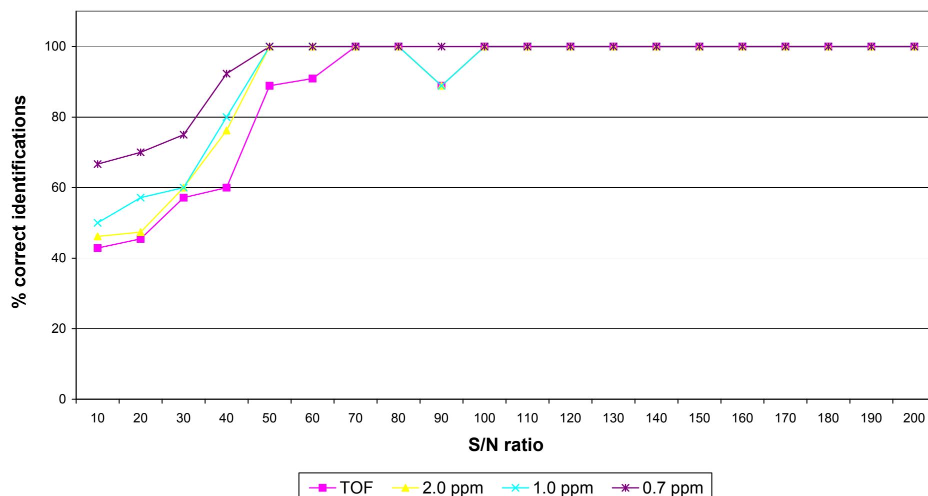


Figure 3. Percentage of correct identifications for MALDI-FT-ICR mass accuracy determined peptide identifications, sorted by S/N ratio, in which a correct identification is defined by identifying a specific peptide peak mass as the same sequence of amino acids, by both MALDI-TOF/TOF (with MALDI-FT-ICR accurate masses for the parent mass) and ESI-Orbitrap.

2.6% of all positive identifications, using a 0.7 ppm mass accuracy window set by MALDI-FT-ICR). In addition, it is very important to note that the false positive matches in sequence were almost exclusively found among the peptides with very low S/N ratios. While this is encouraging, it is important to note that most biologically relevant peptides and proteins in biomarker research are not present in high concentrations [13, 14], and these markers are often masked by the high abundant proteins in the analyzed bio-fluid. Because of this masking, also caused by ion suppression, these low abundant markers result in peaks with low S/N ratios in mass spectra, putting them right in the middle of the danger zone for false identification. An approach employing multiple different and repetitive identification techniques clearly limits the risk of reporting false identifications.

Adding the data of the mass accuracy determinations decreases the number of false positive matches in sequence: the more stringent the window for the accurate mass determination in the MALDI-FT-ICR, the more reliable the results of the database search become. The fact still remains that peptide peaks with lower S/N ratios are more likely to be falsely identified than peptide peaks with high S/N ratios, although the effect is significantly reduced by lowering the mass accuracy window. To obtain the mass accuracy values for lower than 0.7 ppm mass accuracy by MALDI-FT-ICR multiple measurements must be averaged, using the best (internal) calibrations possible, but this would be more time consuming than the current procedure of 0.7 ppm window mass accuracy.

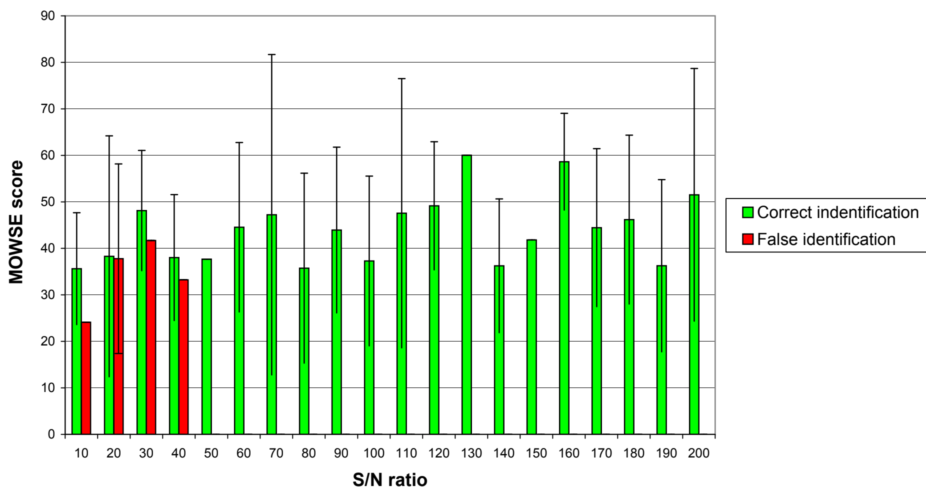


Figure 4. Average Mascot scores of correct and false identifications ordered by S/N ratio, based on 0.7 ppm mass accuracy for the parent ion in the MALDI-FT-ICR. False positive matches in sequence only occurred in the lower S/N ratio range. A correct identification is defined by identifying a specific peptide peak with an identical amino acid sequence by both MALDI-TOF/TOF and ESI-Orbitrap. Standard deviations are shown for columns with three or more identified peptides.

Although the height of the Mascot score is indicative for the reliability of the identification we were unable to correlate this value to false positive matches in sequence in a statistically significant manner when looking at the data separated in subgroups by S/N ratio. Next, we combined all data, regardless of S/N ratio of both the true positive as well as the false positive matches in sequence, of the peptides of which the accurate mass was confirmed with an accuracy of lower than 0.7 ppm by MALDI-FT-ICR. A two-tailed t-test revealed that the correlation between the true positive and false positive matches in sequence had a p-value of 0.0107, which is to say if all data are combined, a clear correlation between the height of the Mascot score and the true or false positive matches in sequence is obtained. The difference in degree of correlation observed between these comparisons is undeniably due to the higher Mascot scores of the correctly identified high abundant peptide peaks, which had larger S/N ratios. Another factor that influences the process of identification is the precursor isolation window. The larger the window is, the bigger the change that other peptide peaks are included in this window and subsequently influence the MS/MS spectrum, resulting in incorrect identifications. The precursor isolation windows used in this experiment are optimized settings for these mass spectrometers. Although it cannot be discounted that these settings may have caused an incorrect identification to occur, it is not likely that this is the cause of any false identification.

As it is often observed, peptides identified by peaks with S/N ratio of 200 or higher were present in the majority, if not all, of the samples. Peptides identified by peaks with low S/N ratios were found in only a few, and often just one, sample. This is because proteins identified by peptide peaks having large S/N ratios are often high abundant proteins in a sample. Indeed, of the eight peptide peaks that were determined to be false positive matches in sequence when a MALDI-FT-ICR window of 0.7 ppm was used, six were peptide peaks found only in a single sample. The other two peptide peaks were identical peak masses found in only two samples in which they were identified as false positive matches in sequence. Because peptide profiling is focused on finding differentially abundant peptide masses between groups, peptide masses that are only present in a single sample are not assigned as being significant when the groups are compared by advanced statistics.

The overlap in proteins identified by different types of mass spectrometers in a single type of cell organelle, has been shown to be 63% [15], which is surprisingly much higher than the 28% we report here. Peptide profiling by MALDI-TOF and SELDI-TOF is very wide-spread within the proteomics research community for biomarker discovery [16-19]. While the determination of peptide masses which are differentially abundant between diseased and control individuals is an important first step, the second step of identifying the differentially abundant peptide peaks is at least equally important to the process. In this process the choice of database is very important as it stands to reason that different databases could potentially lead to different identified peptides and proteins. After identification, a correctly identified protein [20-22] can be immediately assessed for clinical relevance to the disease under investigation or to the biological matrix used. For any conclusion to be drawn from such data it is essential that there is a consensus or guideline on the degree of certainty of the correctness of the identifications. It is known from literature that the reliability of peptide profiling is greatly enhanced by increasing the number of samples measured and by the high mass accuracy offered by modern mass spectrometers. In the present paper we show that the reliability of the second phase (i.e. identification) benefits greatly from using different mass spectrometry techniques in concert.

CONCLUSIONS AND OUTLOOK

It becomes increasingly necessary to develop more robust standards for peptide and protein identification as this is vital for the interpretation of published proteomics results [23]. Thus it may be prudent to add more information to the

identification process, such as: accurate mass of the parent ion; incorporation of sequence information using sophisticated software applications [10]; more rigorous software-based cut-off values for ion scores [24]; using a combination of multiple mass spectrometry techniques to identify specific targeted peptide masses. Another approach would be to combine two or more of these options in the same analysis to optimize the reliability of the result of the MS/MS identification experiments.

Recently developed automation of MALDI-FT-ICR will improve MALDI-based peptide profiling, as this will allow for the measurement of far greater numbers of samples. This will yield far more statistical power to define the differentially abundant peptides at sub ppm mass accuracy. Recent developments showing excellent correlation of peptide concentration to peak intensity accuracy of MALDI-FT-ICR measurements [25] as well as the unparalleled mass accuracy of this device are powerful tools in the ongoing search for biomarkers using proteomics.

A new method of metabolomics profiling, using direct nano-LC-ESI-Orbitrap measurements for profiling and identification in a single experiment could also be very helpful in this field of research. So far the literature shows only metabolomics papers [26-28] using this method.

In conclusion, the reliability of identification by MS/MS of a specific peptide mass observed in a peptide profile is determined by the following four parameters: the S/N ratio of the parent ion, the experimental mass accuracy, the Mascot score, and the frequency of observation in the different samples measured.

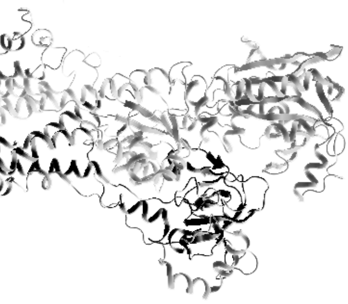
ACKNOWLEDGEMENTS

This work was supported by a program grant provided by the Dutch Multiple Sclerosis Research Foundation to the Rotterdam Multiple Sclerosis research Center ErasMS, the Netherlands Proteomics Centre, the Virgo consortium and the Top Institute Pharma grant D4-102-1.

REFERENCES

1. Pusch, W., et al., *Application of MALDI-TOF mass spectrometry in screening and diagnostic research.* Curr Pharm Des, 2005. **11**: p. 2577-91.
2. Favreau, P., et al., *Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen.* Toxicon, 2006. **47**: p. 676-87.
3. Medzihradszky, K.F., *Peptide sequence analysis.* Methods Enzymol, 2005. **402**: p. 209-44.
4. Hu, S., et al., *Human body fluid proteome analysis.* Proteomics, 2006. **6**: p. 6326-53.
5. Biroccio, A., et al., *Differential post-translational modifications of transthyretin in Alzheimer's disease: a study of the cerebral spinal fluid.* Proteomics, 2006. **6**: p. 2305-13.
6. Cheng, A.J., et al., *Oral cancer plasma tumor marker identified with bead-based affinity-fractionated proteomic technology.* Clin Chem, 2005. **51**: p. 2236-44.
7. Nagalla, S.R., et al., *Proteomic analysis of maternal serum in down syndrome: identification of novel protein biomarkers.* J Proteome Res, 2007. **6**: p. 1245-57.
8. Stoop, M.P., et al., *Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry.* Proteomics, 2008. **8**: p. 1576-85.
9. Nesvizhskii, A.I., *Protein identification by tandem mass spectrometry and sequence database searching.* Methods Mol Biol, 2007. **367**: p. 87-119.
10. Khatun, J., et al., *Incorporating sequence information into the scoring function: a hidden Markov model for improved peptide identification.* Bioinformatics, 2008. **24**: p. 674-81.
11. Dekker, L.J., et al., *FTMS and TOF/TOF mass spectrometry in concert: identifying peptides with high reliability using matrix prespotted MALDI target plates.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **847**: p. 62-4.
12. Titulaer, M.K., et al., *A software application for comparing large numbers of high resolution MALDI-FTICR MS spectra demonstrated by searching candidate biomarkers for glioma blood vessel formation.* BMC Bioinformatics, 2008. **9**: p. 133.
13. Pieper, R., et al., *Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome.* Proteomics, 2003. **3**: p. 422-32.
14. Righetti, P.G., et al., *Proteome analysis in the clinical chemistry laboratory: myth or reality?* Clin Chim Acta, 2005. **357**: p. 123-39.
15. Bodnar, W.M., et al., *Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage.* J Am Soc Mass Spectrom, 2003. **14**: p. 971-9.
16. Belluco, C., et al., *Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage 1 breast cancer.* Ann Surg Oncol, 2007. **14**: p. 2470-6.
17. Nanni, P., et al., *Serum protein profiling in patients with inflammatory bowel diseases using selective solid-phase bulk extraction, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and chemometric data analysis*. Rapid Commun Mass Spectrom, 2007. **21**: p. 4142-8.
18. Villanueva, J., et al., *Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry.* Anal Chem, 2004. **76**: p. 1560-70.
19. Zheng, G.X., et al., *Establishment of serum protein pattern for screening colorectal cancer using SELDI-TOF-MS.* Exp Oncol, 2006. **28**: p. 282-7.

20. Chuthapisith, S., et al., *Proteomic profiling of MCF-7 breast cancer cells with chemoresistance to different types of anti-cancer drugs*. Int J Oncol, 2007. **30**: p. 1545-51.
21. Jeong, H.C., et al., *Proteomic analysis of peripheral T-lymphocytes in patients with asthma.* Chest, 2007. **132**: p. 489-96.
22. Zhang, D., et al., *Proteomic identification of down-regulation of oncoprotein DJ-1 and proteasome activator subunit 1 in hepatitis B virus-infected well-differentiated hepatocellular carcinoma*. Int J Oncol, 2007. **31**: p. 577-84.
23. Tabb, D.L., *What's driving false discovery rates?* J Proteome Res, 2008. **7**: p. 45-6.
24. Vasilescu, J., et al., *Systematic determination of ion score cutoffs based on calculated false positive rates: application for identifying ubiquitinated proteins by tandem mass spectrometry*. J Mass Spectrom, 2008. **43**: p. 296-304.
25. Stoop, M.P., et al., *Quantitative peptide profiling by MALDI-FT-ICR of cerebrospinal fluid obtained from multiple sclerosis patients.* (submitted for publication).
26. Ding, J., et al., *Capillary LC coupled with high-mass measurement accuracy mass spectrometry for metabolic profiling*. Anal Chem, 2007. **79**: p. 6081-93.
27. Dunn, W.B., et al., *Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system.* J Chromatogr B Analyt Technol Biomed Life Sci, 2008.
28. Kamleh, A., et al., *Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules.* Rapid Commun Mass Spectrom, 2008. **22**: p. 1912-1918.



Chapter 4

Multiple sclerosis related proteins identified in CSF by advanced mass spectrometry

Marcel P. Stoop, Lennard J. Dekker, Mark K. Titulaer, Peter C. Burgers, Peter A. E. Sillevis Smitt, Theo M. Luider, and Rogier Q. Hintzen

Proteomics (2008), 8:1576-1585.

ABSTRACT

A total of 164 CSF samples taken from neurological patients were classed into four groups according to the clinical diagnosis: multiple sclerosis (MScl, n=44), clinically isolated syndrome of demyelination (CIS, n=40), other inflammatory neurological disease (OIND, n=26) and other neurological disease (OND, n=54). After tryptic digestion, the samples were measured by MALDI-TOF MS. Spectra were analysed using the R-project software package, in which a peak detection algorithm was developed. Subsequently the peak lists were compared based on ranked data (non-parametric). Significant differences were observed in the comparisons of MScl vs. OND and CIS vs. OND. The comparisons of MScl vs. OIND, and CIS vs. OIND showed fewer significant differences. No significant differences were found in comparisons MScl vs. CIS and OIND vs. OND. MScl and CIS had strikingly similar profiles, probably a reflection of common pathological mechanisms. Three differentially expressed proteins in the comparison of MScl vs. OND were identified: chromogranin A, a potential marker for neurodegeneration; and two important factors in complement mediated inflammatory reaction, clusterin and complement C3. CSF chromogranin A levels were confirmed to be significantly elevated in the MScl group using an ELISA.

INTRODUCTION

Multiple sclerosis (MScl) is a complex disease of the central nervous system (CNS) with presumed autoimmune origin [1]. Its pathology is characterized by a combination of inflammation, demyelination and axonal damage [2]. These processes are not uniformly represented across patient populations but can predominate selectively in individual patients, contributing to the heterogeneity of phenotypic expression of the disease [3].

Due to the complex nature of the disease, there is probably no single protein or peptide that can serve as a biomarker for MScl in a clinically relevant way. However, individuals with MScl might be differentiated from healthy individuals by a proteomic pattern consisting of a set of distinct individual proteins or peptides that are not independently reliable indicators of disease state. Because the disease process in MScl is located in the CNS, cerebrospinal fluid (CSF) is a promising body fluid in which to search for biomarkers and disease-associated proteins and peptides [4, 5]. The absence of an active clotting system in CSF compared to serum makes it an attractive bio-fluid to perform proteomic studies. Biomarkers for MScl could potentially identify relevant biological pathways for MScl, which could help clarify currently unsolved issues in the pathology and etiology of the disease [6]. Additionally, biomarkers could also be useful as a new method for diagnosis of MScl, which might enable earlier detection of the disease, or as a tool to identify clinically relevant subgroups. Earlier detection, and subsequently early treatment, will retard the long-term evolution of the disease [7].

Recently, a study was published in which a cleavage product of the protein Cystatin C was identified as a potential biomarker for a subgroup of MScl [8]. However, a later report suggests that this cleavage product is merely a freezing artifact caused by freeze-thaw cycles [9]. Further research is required to clarify this. Other studies used pooled CSF samples of MScl patients after ultrafiltration [10] or 2D gel electrophoresis [11] as sample preparation steps in their efforts to identify proteins in human CSF.

The objective of this study was to examine CSF peptide profiles of MScl patients in order to identify the peptides and proteins that differ between the patient groups. Using MALDI-TOF Mass Spectrometry we analyzed a total of 164 CSF samples. After tryptic digestion of the CSF proteins, the peptide mixtures were measured by MALDI-TOF MS and subsequently analyzed using a bio-informatics tool [4]. MALDI Mass spectrometry was used in this study because it is uniquely suited for high-throughput peptide profiling, as it is a relatively fast and accurate way of analyzing complex protein and peptide mixtures. Although it is also a useful tool for the identification of peptides and proteins, we also confirmed the iden-

tifications by more advanced mass spectrometry techniques (MALDI-FTMS and ESI-Orbitrap). The ESI-orbitrap also identified an additional eight peptides that were not identified by the MALDI-TOF/TOF experiments.

MATERIALS AND METHODS

Patient selection

All samples analyzed in this study were taken from patients that were followed prospectively by the Rotterdam Multiple Sclerosis Center and the department of Neurology at Erasmus University Medical Center (Rotterdam, the Netherlands). The samples were all specifically selected from untreated patients. These samples were selected and classified into four groups by an experienced neurologist (RQH). The first designated group consisted of CSF samples from patients suffering from Multiple Sclerosis (MScl, n=44). The second group consisted of samples from patients who were diagnosed with a clinically isolated syndrome of demyelination (CIS, n=40), which is considered as a pre-stage of MScl. The third group of samples was taken from patients with another inflammatory neurological disease (OIND, n=26), to be used as controls. The fourth group, which was also a control group, consisted of samples from patients with some other neurological disease (OND, n=54). All these diagnoses were based on data derived from the examination at the time the CSF sample was taken.

Immediately after sampling, the CSF samples were centrifuged to discard cells and cellular elements (10 minutes at 3000 rpm). The samples were subsequently used for routine CSF diagnostics. This included quantification of total protein and high abundant protein (albumin and immunoglobulin G) concentration, assessment of the number of oligoclonal bands, which are present in most, but not all MScl patients, as well as quantification of the intrathecal cell count. The remaining volume of the samples was aliquoted and stored at -80°C, where they remained until sample preparation for this study.

Sample preparation

Prior to all sample preparation procedures, all samples were blinded. They were subsequently measured and analyzed in a random order. Twenty µl of each CSF sample was put into a 96-microtiter well plate (Nunc. Low binding, VWR, the Netherlands), and an equal amount of 0.2% Rapigest (Waters, USA) in 50 mM ammonium bicarbonate buffer was added to each well. Following a two-minute incubation period at 37°C, 4 µl 0.1 µg/µl gold grade trypsin (Promega, USA) / 3 mM Tris HCl (pH 8.0) was added to each well. The samples were incubated at 37°C for

two hours. To obtain a final concentration of 30-50 mM HCl (pH < 2), 2 µl of 500 mM HCl was added. The samples were then incubated for 45 minutes at 37°C, which stopped the digestion reaction.

Following the digestion procedure, the samples were desalted using 96-well zip C18 micro titer plates (Millipore, USA), which had been pre-wetted and washed twice with 100 µl acetonitrile per well. The samples were centrifuged (Multifuge 3 S-R, Goffin Meyvis, the Netherlands) at 2000 rpm for five minutes. After the washing step, 3 µl acetonitrile were put on the C18 resin to prevent drying. Each trypsin digested CSF sample was mixed with 200 µl HPLC grade water / 0.1% TFA. The samples were subsequently put onto the washed and pre-wetted 96-well zip C18 plate, and 30 minutes of centrifugation at 1500 rpm was used to load the peptides onto the C18 material. The wells were washed twice with 100 µl 0.1% TFA (5 minutes, 1750 rpm). An elution volume of 15 µl 50% acetonitrile/ 0.1% TFA was used to elute the samples in a new 96-well plate. The centrifugation period for the elution step was 30 minutes at 1600 rpm. After this elution step, the samples were stored in 96-well plates, which were covered with aluminum seals, at 4°C.

Measurement

A matrix solution was made by dissolving 2 mg α-cyano-4 hydroxy-cinnamic acid (HCCA) in 1 ml acetonitrile, using an ultrasonic bath for 30 minutes. Two µl of elute of each sample was mixed with eight µl of the matrix solution, 0.5 µl of which was spotted onto a MALDI target (600/384 AnchorChip™ with transponder plate, Bruker Daltonics, Germany). All samples were spotted in duplicate. The digestion step and the subsequent measurement in duplicate were performed three times for each sample, resulting in 6 spectra per sample. The samples were all measured using the automated measurement feature of the MALDI-TOF MS (Ultraflex, Bruker Daltonics, Germany). The standard method for peptide measurements (Proteomics_HPC) was used on the MALDI-TOF MS, with the measurement range set to 500-3400 Dalton (Da).

The following settings were used for the automated measurements: the initial laser power was 25%, and the maximum was 45%. The peak with the highest intensity above the 750 Da had to have a signal-to-noise ratio of at least 5 and a minimum resolution of 5000. Every 50 laser shots the sum spectrum was checked for these criteria. It was rejected if it did not meet these criteria. If 20 sum spectra of 50 laser shots met these criteria, they were combined and saved. If 25 consecutive sum spectra of 50 shots had been rejected, the measurement of the spot was ended and the measurement of the next spot commenced.

Analysis

The raw binary data files of the MALDI-TOF MS experiments were first converted to ASCII files containing the measured intensities for all channel indices of the spectra. All spectrum files were designated a group number (1 – MScl, 2 – CIS, 3 – OIND, 4 – OND) and serial numbered. To calibrate the channel numbers to masses we used a quadratic fit with a number of internal calibrants, which were five omnipresent tryptic albumin peptides (927.4934, 1226.6051, 1467.8430, 1875.0156 and 2045.0953). The algorithm described by Dekker and co-workers [12, 13] was used for peak detection, performed in the statistical language R (http://www.r-project.org). A percentile threshold of 96% was chosen, meaning that the intensity of the peak position must belong to the 4% highest intensity values of the spectrum. The mass window (minimum distance between two adjacent peak positions) was set at 0.5 Da. After the generation of a peak list for each spectrum using the peak detection algorithm, an analyses matrix was created in R. During this process all peak lists of every sample were checked for the presence of all peaks. Only absence or presence of the peaks was scored. Thus, a peak position that was present in all peak lists of one sample, was designated with a 6 and a peak position present in half of the peak lists was designated with a 3. In this matrix, all samples were tagged with their group number and with a number ranging from 0 to 6 for all separate peak positions. The matrix was subsequently used for statistical analysis of the data. Using a univariate analysis in R, a p-value was determined for every peak position. The Wilcoxon-Mann-Whitney test was used for comparison between the groups [14, 15]. For data that does not follow a known theoretical distribution, one can use non-parametric tests. One such test is the Wilcoxon-Mann-Whitney test. Data from multiple groups is combined and the data are tagged so that the group to which they belong is recorded. The combined data set is then sorted according to the value of the data. This type of statistical test is very well suited for the comparison of pre-defined groups, as was the case in these experiments. A cross validation was performed on the same data by randomly assigning a group number to each CSF sample and then repeating the Wilcoxon-Mann-Whitney test. This scrambling procedure was subsequently repeated 10.000 times, which gives an impression of the probability to find a significant differentially expressed peptide by chance.

Identification

The next step in the process was identification of the differentially expressed peptides. Due to the limitations of the MALDI-TOF strategy, the differentially expressed peptides could not be identified immediately from the complete CSF sample, due to the high amount of peaks per spectrum. So, for identification of the peptides alternative methods were needed. We used two different strategies. Firstly, we

determined the accurate masses of the peptides with a 9.4 Tesla (MALDI) Fourier Transform Mass Spectrometer (FTMS, Bruker Daltonics, Germany). Due to the excellent mass accuracy and precision of this technique it is suited to determine the masses of the peptides up to an accuracy of better than 1 ppm. Because the HCCA matrix is not compatible with orthogonal MALDI (Apex I, Bruker Daltonics, Germany), we used another matrix molecule, 2,5-dihydroxy benzoic acid (DHB).

Secondly, we used a nano-LC system (Ultimate system, Dionex, USA) to separate the digested CSF peptides on a C18 reversed phase column, followed by off-line MALDI-TOF/TOF mass spectrometry. The digested CSF sample was injected and separated on this column during a 130 min LC run, using a 90 minute gradient of increasing concentrations of acetonitrile (Solvent A: 95% TFA (0.1%) in water, 5% ACN (v/v); Solvent B: 20% TFA (0.1%) in water, 80% ACN), and subsequently spotted on a MALDI target plate. The MALDI target plate used in this nano-LC experiment was a pre-spotted target plate (PAC 384 plate, Bruker Daltonics, Germany). On this plate the matrix was already previously applied and only the sample has to be added. By dividing a single sample into 384 separate spots and corresponding spectra, it proved to be possible to acquire the required mass windows for several of the differentially expressed peptides. Measurement of the pre-spotted plate was done in an automated way, using the Warp-LC software package, version 1.0 (Bruker Daltonics, Germany) on the MALDI-TOF/TOF, which first generated a mass spectrum for all spots and then proceeded to generate MS/MS spectra, in a data dependent manner, for peaks which were suitable for MS/MS by a series of pre-set conditions, such as, for example, a sufficiently high signal-to-noise ratio. Identifications were obtained by database searches of the MS/MS spectra using Mascot search program, version 2.1 and the SwissProt database (version 51.2), using the standard settings of the Warp-LC software package. The score threshold for peptide identification was a Mascot score of over 54, which correlates with a p-value < 0.05. Identifications were confirmed by determination of accurate mass by MALDI-FTMS.

In a separate effort to identify the differentially expressed peptides we also used an Orbitrap mass spectrometer with electrospray ionization (Thermo Electron, Germany), using a C18 column nano-LC system online connected to the device. Five microliter of the digested CSF was injected into the nano-LC and subsequently the digested peptides were separated on the C18 column. After ionization the peptides were measured in the Orbitrap, using a data-dependent acquisition mode for the MS/MS identification step, which occurred in the ion trap. The MS/MS identifications were obtained using the Bioworks 3.2 software package (Thermo Fisher Scientific, Germany) and its SEQUEST feature, using the standard settings, with XC scores of 1.5, 2.0 and 2.5 for single, double and triple

charged ions respectively. With this method we also used the SwissProt database for database searching as mentioned above.

Immunoassay

As specified by the manufacturer (DakoCytomation, Denmark), we used a commercially available ELISA kit for Chromogranin A to determine the chromogranin A concentrations in the CSF samples. The sample volumes available were not sufficient to do this ELISA for all samples. 123 samples from the original sample set were evaluated (32 MScl, 32 CIS, 25 OIND and 34 OND samples). The chromogranin A concentrations in CSF of MScl patients and controls were also determined for external validation purposes. To this end we used a separate group of MScl (n = 19) and control (OND) samples (n = 18) which included among others non neurological cancer, intracranial hypertension, Sneddon syndrome and headache.

RESULTS

Clinical information

Of the 44 patients in the MScl group, 30 patients had relapsing remitting (RR) MScl and 14 had primary progressive (PP) MScl (Table 1). The mean age of the patients in this group was 42.1 ± 11.3 years, and the mean protein concentration of the CSF was 0.41 ± 0.13 g/l. The group contained 12 males and 32 females. The median of the time that these patients had been afflicted with MScl was nearly two years.

The CIS group (n=40) consisted of 12 males and 28 females, nearly 70 percent of whom had been diagnosed with optic neuritis (Table 1). The other patients were diagnosed with one, or a combination, of the other possible localizations of a CIS, like for example brainstem syndromes or myelitis [16]. The mean age of this second group was 33.7 ± 9.4 years, and the mean protein concentration of the CSF was 0.36 ± 0.11 g/l. The median time between the sampling of the CSF and the occurrence of the first symptoms in these patients was three months.

The OIND group (n=26) contained 7 males and 19 females, with a mean age of 49.6 ± 16.3 years. The mean CSF protein concentration was 0.43 ± 0.16 g/l (Table 1). Diagnoses of the patients in this group were bacterial and sterile infections (n=17), vasculitis (n=7), and Guillain Barré Syndrome (n=2).

The 54 patients in the OND group had a mean age of 48.1 ± 16.7 years. This group contained 23 males and 31 females, and the CSF samples had a mean protein concentration of 0.42 ± 0.22 g/l (Table 1). The diagnoses set for these patients were headaches (n=19), neurological degeneration (n=11), non-neurological disease (n=10), infarct (n=8), brain tumor (n=4), and other (n=2)

Table 1. Clinical information, including routine protein quantifications, of all patients included in this study.

	MScl (n=44)	**CIS** (n=40)	**OIND** (n=26)	**OND** (n=54)
Males/Females	12/32	12/28	7/19	23/31
% of samples oligoclonal positive	72.7%	60.0%	7.7%	3.7%
Median time between first MScl or CIS related complaint and sampling in months (range)	22 (0-232)	3 (0-72)#		
Protein concentration (g/l). SD in brackets	0.406 (0.135)	0.358 (0.106)	0.435 (0.161)	0.422 (0.217)
Albumin concentration (g/l). SD in brackets	0.239 (0.097)	0.210 (0.080)	0.250 (0.160)	0.280 (0.190)
IgG concentration (g/l). SD in brackets	0.060 (0.033)	0.048 (0.029)	0.055 (0.045)	0.043 (0.032)
Diagnoses	- 30 RR MScl - 14 PP MScl	- 25 optic neuritis CIS - 7 myelitis CIS - 5 brainstem CIS - 3 other CIS	- 17 infections (bacterial and sterile) - 7 vasculitis - 2 Guillain Barré Syndrome	- 19 isolated headaches - 11 neurological degeneration - 10 non neurological disease - 8 infarct - 4 brain tumor - 2 other

Excepting two patients (72 and 27 months), all CIS samples were taken within twenty months after the first symptoms were observed.

- Concentrations: $p>0.05$ for all comparisons (two-tailed t-test), except total protein concentration for CIS-OIND ($p=0.026$) and IgG concentration for MScl-OND ($p=0.035$).

The total protein concentrations of the CSF samples did not differ significantly between the groups when subjected to a two-tailed t-test. Also, the albumin and immunoglobulin G (IgG) concentrations of the CSF samples did not differ significantly between the groups.

Peak detection and data analysis

An average of 515 peaks was detected per MALDI-TOF spectrum. After spectrum conversion, the matrix was created, which consisted of a total of 1755 peaks. The significance of difference in distribution over the four groups was tested for each peak.

The significant difference between the MScl group and the OND group ($p<0.05$: 132 peaks; $p<0.01$: 44 peaks) indicates obvious differences in the CSF of MScl patients compared to that of the non-inflammatory control patients. The obvious skew towards the low p-values is clearly visible in the p-value histogram (figure 1A). The largest statistical difference (figure 1B) was observed in the comparison of the CIS group and the OND group ($p<0.05$: 161 peaks; $p<0.01$: 60 peaks).

The comparison of the MScl group with the OIND group (figure 1C), and also that of the CIS group with the OIND group (figure 1D), shows a number of peak positions with a statistically significant p-value ($p<0.05$: 127 peaks; $p<0.01$: 12 peaks, and $p<0.05$: 134 peaks; $p<0.01$: 27 peaks, respectively). Both these p-value histograms are also skewed towards the lower p-values, and, in this region, the height of the bars clearly exceeds the height of the red line, which reflects the expected statistical background.

Comparison of the MScl group with the CIS group (figure 1E) shows few peaks with a significant p-value ($p<0.05$: 50 peaks; $p<0.01$: 10 peaks). The height of the bars in the low p-value area clearly lies below the red line generated by the randomized comparison, indicating that there is no significant difference between the CSF samples of the patients classified in the MScl group compared to the patients in the CIS group. The comparison between the two control groups (OIND vs. OND, figure 1F) shows only a slightly higher number of statistically significant peak positions ($p<0.05$: 101 peaks; $p<0.01$: 12 peaks). The height of the bars in the low p-value area lies at about the same height as the red line generated by the randomized comparison, which indicates that these two groups are not statistically significantly different.

Identification

Using the combination of the FTMS and the nano-LC MALDI-TOF/TOF techniques we were able to determine the amino acid sequence of three of the differentially expressed peptides in the comparison between MScl and OND. These peptides were identified to be tryptic peptides of chromogranin A, clusterin and complement C3 by offline nano-LC MALDI-TOF/TOF. Determination of the accurate mass of these three tryptic peptides by FTMS resulted in parent ion masses differing 0.19, 0.27 and 0.67 ppm from the theoretical masses for those specific peptides of chromogranin A, clusterin and complement C3, respectively. Although the Mascot

Figure 1. Graphical representation of the results of the Wilcoxon-Mann-Whitney tests for the comparisons of the four groups. The heights of the bars represent the total number of peaks with a specific p-value interval, whereas the green bar indicates the number of peaks that were predominantly present in the first mentioned group, and the blue bar indicates the number of peaks predominantly present in the second mentioned group. The number of peaks found for each p-value after randomization is represented by the red line, which can be viewed as a baseline. Any low p-value bar that exceeds the height of the baseline indicates statistically significant differences in that particular comparison. Figure 1A shows the comparison of MScl with OND, in which the large skewing to the left of the p-value range indicates that these groups differ significantly. The same pattern can be seen in figure 1B, which represents the comparison of CIS with OND. Smaller, but still significant differences can be observed in the comparison of MScl with OIND and the comparison of CIS with OIND, in figures 1C and 1D, respectively. The comparison of MScl with CIS (figure 1E) and the comparison of OIND with OND (figure 1F) do not show significant differences.

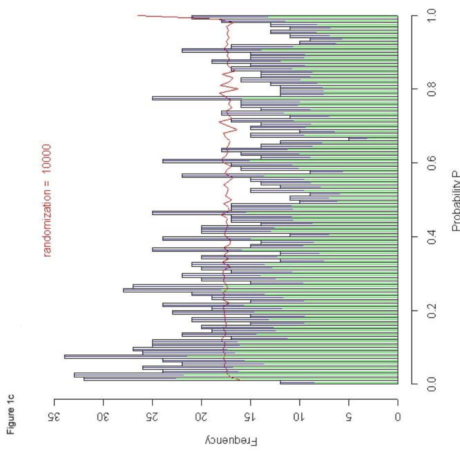


Figure 1c
randomization = 10000
Frequency
Probability P

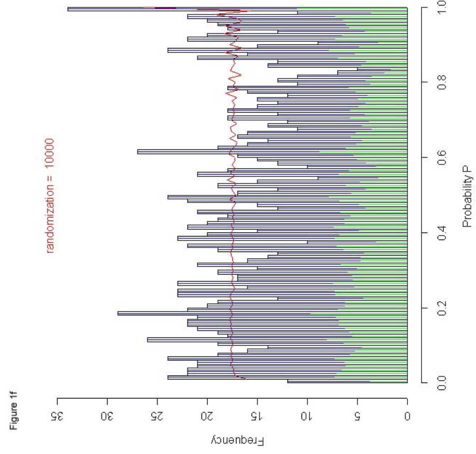


Figure 1f
randomization = 10000
Frequency
Probability P

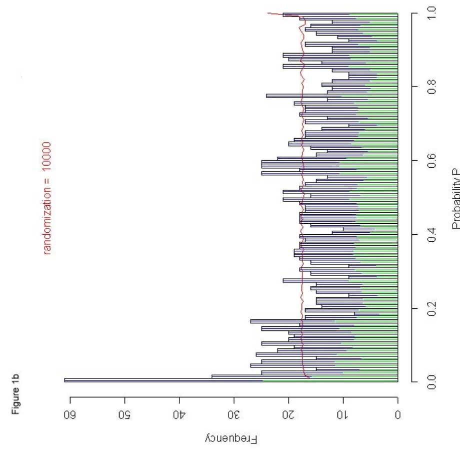


Figure 1b
randomization = 10000
Frequency
Probability P

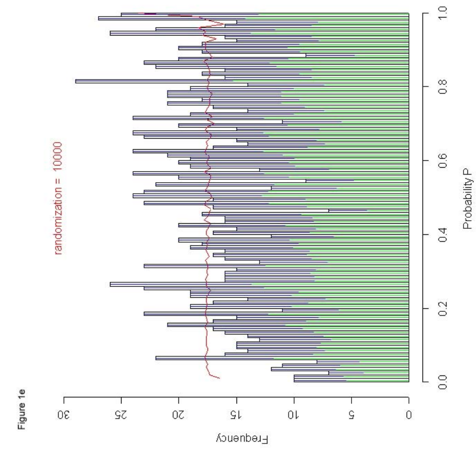


Figure 1e
randomization = 10000
Frequency
Probability P

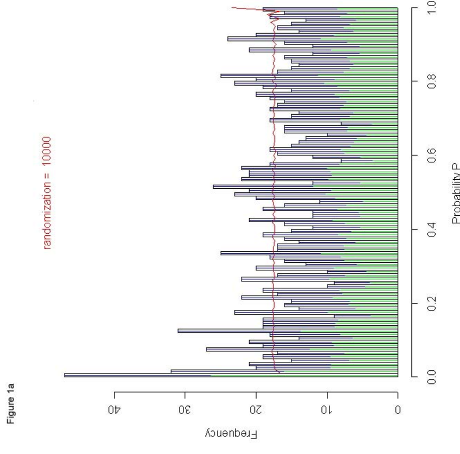


Figure 1a
randomization = 10000
Frequency
Probability P

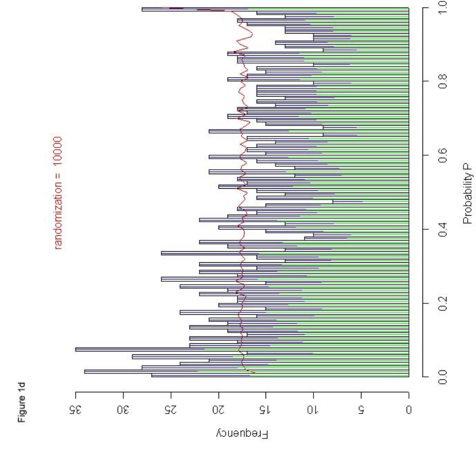


Figure 1d
randomization = 10000
Frequency
Probability P

Table 2. Identified differentially expressed peptides and corresponding proteins (p<0.01) amongst the patient groups.

					MALDI-FTMS*		ESI-Orbitrap		
Comparison	Protein	Accession number	Peptide sequence	Theoretical mass	Measured mass	Δ (ppm)	Measured mass	Δ (ppm)	No. of pept.
MScl-OND	Chromogranin A	P10645	YPGPQAEGDSE GLSQGLVDR	2074.9675	2074.9679	0.193	2074.970	1.205	5
MScl-OND	Clusterin	P10909	TLLSNLEEAK	1117.6099	1117.6102	0.268	1117.610	0.089	9
MScl-OND	Complement C3	P01024	AGDFLEANYMN LQR	1641.7689	1641.7678	0.670	-	-	-
MScl-OIND	Complement C4B	P0C0L5	ASAGLLGAHAA AITAYALTLTK	2085.1703	2085.1684	0.911	2085.173	1.295	1
CIS-**OND**	Beta V spectrin	Q9NRC6	WINNVFQCGQA GIKIR	1903.9959	1903.9944	0.193	1904.000	2.153	1
CIS-**OND**	Hypothetical protein XP_011125	-	ISHELDSASSEVN	1387.6335	1387.6348	0.937	1387.635	1.081	2
CIS-OIND	Apolipoprotein D	P05090	VLNQELR	871.4996	871.4991	0.574	871.4994	0.229	1
CIS-OIND	Complement C4A	P0C0L4	GLQDEDGYR	1052.4643	1052.4641	0.190	1052.464	0.285	10
CIS-OIND	Contactin 1	Q12860	VQVTSQEYSAR	1267.6277	1267.6288	0.868	1267.628	0.237	4
CIS-OIND	Neuronal pentraxin receptor	O95502	QTALQQEAR	1044.5432	1044.5435	0.287	1044.543	0.191	2
CIS-**OIND**	RNA binding motif protein 7	Q9Y580	VTEELLFELFHQA GPVIKVK	2297.2903	2297.2881	0.958	2297.286	1.871	1

* Number of peptides in the direct (without nano-LC fractionation) MALDI comparison was 1. Column 1: Bold denotes group with elevated expression for that specific protein, 1 missed cleavage allowed.

score for the clusterin peptide is below 54, the additional accurate mass obtained by FTMS and the XC score (3.22, doubly charged) obtained for this peptide validated this identification in separate ways.

Using the sequencing capabilities of the Orbitrap mass spectrometer we were able to identify 10 peptides from the list of differentiating peptide peaks (Table 2 and 3). The accurate masses of the parent ions of these identified peptides were further confirmed by FTMS. All identified peptides showed accuracy below 1 ppm for FTMS (Table 2). The additional peptides identified for each protein using the Orbitrap set-up were not detected using the MALDI-TOF profiling approach and thus they were not present in the comparison matrix.

Testing of chromogranin A levels by ELISA showed that MScl samples had higher chromogranin A concentrations than non inflamed control samples (table 4) which was in agreement with the proteomics analysis. The difference in concentration was statistically significant (p = 0.00034). The external validation in a new set of samples, showed that these MScl samples also contained a higher chromogranin A concentration as compared to the chromogranin A concentration in non inflamed control samples.

Table 3. Sequence coverage and scores of significantly differentially expressed peptides found by MALDI-MS.

				MALDI-TOF/TOF	ESI-Orbitrap	
Protein	**Peptide**	**Sequence coverage* (%)**	**Protein Mol. Weight (Da)**	**Mascot Score**	**SEQUEST XC value**	**Charge**
Chromogranin A	YPGPQAEGDSE GLSQGLVDR	20.3	50730	133	3.51[H]	2
Clusterin	TLLSNLEEAK	22.3	52495	50	3.22[H]	2
Complement C3	AGDFLEANYMN LQR	0.8	187148	75	- [X]	-
Complement C4B	ASAGLLGAHAA AITAYALTLTK	1.3	192739	-	3.84[H]	2
Beta V spectrin	WINNVFQCGQA GIKIR	0.4	416835	-	2.07	2
Hypothetical protein XP_011125	ISHELDSASSEVN	-	-	-	3.42[H]	2
Apolipoprotein D	VLNQELR	3.7	21276	-	2.20	2
Complement C4A	GLQDEDGYR	7.4	192771	-	2.64[H]	2
Contactin 1	VQVTSQEYSAR	3.8	113320	-	3.01[H]	2
Neuronal pentraxin receptor	QTALQQEAR	3.6	52718	-	2.24[X]	2
RNA binding motif protein 7	VTEELLFELFHQA GPVIKVK	7.5	30504	-	2.25	2

- Not found.
* Using SwissProtein database (October 2006).
[H] XC value in agreement with the HUPO criteria (XC = 1.8, 2.5 and 3.1 for single, double and triple charged ions respectively).
[X] Other peptide(s) of this protein were identified with XC values in agreement with the HUPO criteria.

Table 4. Chromogranin A concentration (ELISA results) is significantly higher in CSF of MScl patients as compared to non inflamed controls in both the original sample set (p=0.00034) as well as in the external validation sample tset (p=0.02), using a t-test. T-tests for the other comparisons did not show significant differences in the chromogranin A concentrations. The number of tested samples from the original sample set was 123 (32 MScl, 34 OND, 32 CIS and 25 OIND). External validation was done by testing new 19 MScl and 18 OND samples. Concentrations in U/L.

	MScl	OND	CIS	OIND
Concentration in original sample set	251.85	166.26	213.86	204.23
SD (+/-)	91.18	92.28	89.07	80.54
p-value in original sample set (MScl vs OND)	0.00034			
Concentration in external validation sample set	240.44	139.86		
SD (+/-)	167.53	68.80		
p-value in external validation sample set MScl vs OND)	0.02			

DISCUSSION

A main observation of this study is that, on the basis of proteomic CSF analysis, the group of MScl patients could be clearly distinguished from the non-inflammatory neurological controls. However, as MScl is considered to be an inflammatory disease, it was imperative to include a group of controls covering other CNS inflammatory neurological diseases. We were also able to clearly differentiate between CSF peptide profiles of the MScl patients and inflammatory neurological controls. As MScl has both an inflammatory and a neurodegenerative component [17], comparison with both control groups is essential to ensure that the control groups covered both the inflammatory and the non-inflammatory phase of the disease.

We were also able to make similar distinctions in the proteomic CSF analyses of the CIS patients when compared to the non-inflammatory controls, as well as in the comparison of the CIS patients with the inflammatory neurological controls. The fact that the MScl and CIS groups behaved similarly fits with the clinical observations that, in the majority of the cases, CIS is a pre-stage of MScl [18, 19]. It can be assumed that similar immunopathological and neurodegenerative processes take place in both conditions [20]. Another indication of the similarities between the proteomic profiles of the CSF samples of the patients in the MScl group and those of the CIS group is the overlap of the peptides found in the comparisons of MScl with the non-inflammatory controls (OND) and the comparison of CIS with OND. Eleven of the forty-four peptide peaks (25%) found to be significantly ($p<0.01$) different in the comparison of MScl and OND was also found to be significantly different in the comparison of CIS and OND. If the significance cut-off is set to $p<0.05$, this overlap increases to nearly 38%.

We also compared two major distinct disease courses of MScl, PP MScl and RR MScl. Specifically for this sub-comparison we had included a relatively high number of PP MScl patients in this study (32% of the MScl samples (normal 10-15% of the total MScl population)). However, we did not observe any significant differences between PP MScl and RR MScl (data not shown).

Because pathology changes over time in MScl [21-24] one could imagine differences in CSF proteomic profiles as the disease duration progresses. The MScl samples were divided into a short and a long disease duration group, according to their position in relation to the median time between first symptoms and CSF sampling. No significant differences were observed in this comparison, compared to the statistical background (data not shown). It has been suggested that the course of the disease is more benign in MScl patients without oligoclonal Immunoglobulin G bands in CSF [25]. One might therefore expect to find other differences in CSF composition when comparing MScl patients with oligoclonal Immunoglobulin G

bands in the CSF versus MScl patients without these bands. However, we did not observe this. Finally, also no differences were observed between males and females, compared to the statistical background (data not shown). These sub-analyses lend support to the interpretation that common peptide patterns are shared within the whole group of MScl patients.

It has been shown that factors such as sample stability and a low number of measurements per sample can cause difficulties regarding the reproducibility of proteomic profiling studies [26]. Other studies have shown the low reproducibility of peak height in MALDI-TOF MS [13]. The method described here is less affected by these variations, as the heights of the peaks were not included in this analysis because quantitative measurements of peak heights with MALDI TOF MS are poorly reproducible, with standard deviations up to 30% [13]. In this study only the absence or presence of the peaks was scored.

Considering the fact that these data are a summation of three separate MALDI-TOF experiments, in which CSF samples were measured in duplicate, we also performed three separate analyses for these three experiments. A similar p-value diagram for each experiment was obtained, compared to the combined experiments. Similarly, the differentially expressed peptides that were found in the total analysis also had low p-values in the separate analyses, showing that these results were reproducible inter assay wise.

Another beneficial factor is the lower protein load, especially of high-abundant proteins such as albumin in CSF [27, 28], which allows CSF peptide profiling without complicated sample pre-treatment steps (e.g. high-abundant protein depletion) that are essential to serum peptide profiling. Other studies have already delved into the CSF proteomic profile. Although these studies identified a number of interesting proteins that were present in CSF of MScl patients, these analyses were performed in a setting with limited numbers of patients [29, 30]. It is hard to compare these studies, as the set-up and objectives have been very different. The studies mentioned aimed to identify as many proteins as possible in CSF of MScl patients, whereas we aimed to pinpoint peptide distinctions between MScl patients and controls. The method used in our study also differed from these studies in the fact that our sample pre-treatment was much less complex, as we did not perform a two-dimensional gel electrophoresis procedure prior to the trypsin digestion. Another difference is that the number of samples we analyzed is considerably higher, which allowed for more advanced statistics during the analysis.

In total, we identified eleven proteins in association with MScl (Table 2). Five of them, neuronal pentraxin receptor, contactin 1, beta V spectrin, RNA binding motif protein 7 and hypothetical protein XP_011125, have not been reported in association with MScl previously. These proteins may help to reveal new pathways associated

with MScl. Neuronal pentraxin receptor is a molecule implied in synapse formation and neuronal remodeling [31]. It is also suggested to be involved in a pathway responsible for the transport of taipoxin into synapses and that this may represent a novel neuronal uptake pathway involved in the clearance of synaptic debris. Contactins mediate cell surface interactions during nervous system development. They are involved in the formation of paranodal axo-glial junctions in myelinated peripheral nerves and in the signaling between axons and myelinating glial cells. Beta V spectrin is expressed at very low levels in many tissues, with strongest expression in, among others, cerebellum and spinal cord. It is now well recognized that spectrins may contribute to the establishment and maintenance of membrane order [32].

The other six proteins, apolipoprotein D, chromogranin A, clusterin, complement C3, complement C4A and complement C4B have been previously reported in association with MScl. Apolipoprotein D, which is involved in the removal of lipids during nerve cell degeneration, has been reported previously to be present in elevated levels in MScl [33]. The relevance for the role of complement in MScl pathogenesis is a matter of a long debate, with several papers suggesting elevated levels in affected tissue and in CSF [34, 35]. This detection of several complement components in the CSF of MScl patients lends further support to the role of complement in this disease.

The association of CSF clusterin, also known as apolipoprotein J, with MScl is of particular interest in this respect, because this molecule serves as a regulator (inhibitor) of complement activity. In addition, it is considered as an important factor in neuronal cell survival [36]. There is an earlier report that suggested elevated CSF clusterin levels in patients with central nervous system inflammation [37]. Finally, the detection of chromogranin A in association with MScl is interesting. Although the neurodegenerative component of MScl pathology has remained an enigma [38], an important role of innate immune cells and signals is obvious. It is of interest that chromogranin A can induce neurotoxicity in microglial cells [39]. It will be of interest to determine whether high CSF chromogranin A levels are associated with more aggressive clinical disease types.

In conclusion, the MALDI-TOF analysis of tryptic digested CSF proteins showed significant differences between MScl patients and control patients (MScl vs. OND: 44 peptides with $p<0.01$), as well as significant differences between CIS patients and control patients (CIS vs. OND: 60 peptides with $p<0.01$). Among these proteins were important factors in complement mediated inflammatory reaction, such as complement C3, complement C4A, complement C4B and clusterin, as well as several proteins that were not previously reported to be differentially expressed in MScl and its pre-stage (CIS). These new potential markers include functionally

interesting proteins such as chromogranin A, neuronal pentraxin receptor and contactin 1.

ACKNOWLEDGEMENTS

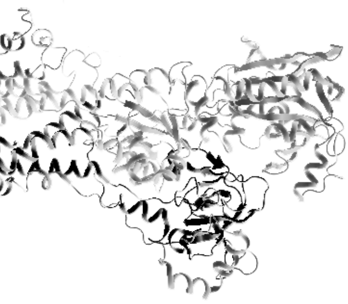
This work was supported by a program grant provided by the Dutch Multiple Sclerosis Research Foundation to the Rotterdam Multiple Sclerosis research Center, ErasMS, the Netherlands Proteomics Centre and the Top Institute Pharma grant D4-102-1.

The authors would like to thank Thermo Fisher Scientific, and in particular Wilfried Voorhorst and Wolfgang Metelmann-Strupat for assistance with the Orbitrap experiments.

REFERENCES

1. Keegan, B.M., et al., *Multiple sclerosis.* Annu Rev Med, 2002. **53**: p. 285-302.
2. Compston, A., et al., *Multiple sclerosis.* Lancet, 2002. **359**: p. 1221-31.
3. Bielekova, B., et al., *Development of biomarkers in multiple sclerosis.* Brain, 2004. **127**: p. 1463-78.
4. Romeo, M.J., et al., *CSF proteome: a protein repository for potential biomarker identification.* Expert Rev Proteomics, 2005. **2**: p. 57-70.
5. Yuan, X., et al., *Proteomics analysis of prefractionated human lumbar cerebrospinal fluid.* Proteomics, 2005. **5**: p. 541-50.
6. Greenstein, J.I., *Current concepts of the cellular and molecular pathophysiology of multiple sclerosis.* Dev Neurobiol, 2007. **67**: p. 1248-65.
7. Tintore, M., *Early MS treatment.* Int MS J, 2007. **14**: p. 5-10.
8. Irani, D.N., et al., *Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis.* Ann Neurol, 2006. **59**: p. 237-47.
9. Del Boccio, P., et al., *Cleavage of cystatin C is not associated with multiple sclerosis.* Ann Neurol, 2007. **62**: p. 201-4; discussion 205.
10. Noben, J.P., et al., *Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography, and mass spectrometry.* J Proteome Res, 2006. **5**: p. 1647-57.
11. Hammack, B.N., et al., *Improved resolution of human cerebrospinal fluid proteins on two-dimensional gels.* Mult Scler, 2003. **9**: p. 472-5.
12. Dekker, L.J., et al., *MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancer.* Mol Cell Proteomics, 2005. **4**: p. 1341-9.
13. Dekker, L.J., et al., *A new method to analyze matrix-assisted laser desorption/ionization time-of-flight peptide profiling mass spectra.* Rapid Commun Mass Spectrom, 2005. **19**: p. 865-70.
14. Gaddis, G.M., et al., *Introduction to biostatistics: Part 5, Statistical inference techniques for hypothesis testing with nonparametric data.* Ann Emerg Med, 1990. **19**: p. 1054-9.
15. Titulaer, M.K., et al., *A database application for pre-processing, storage and comparison of mass spectra derived from patients and controls.* BMC Bioinformatics, 2006. **7**: p. 403.
16. Tintore, M., et al., *Is optic neuritis more benign than other first attacks in multiple sclerosis?* Ann Neurol, 2005. **57**: p. 210-5.
17. Bruck, W., et al., *Inflammation and degeneration in multiple sclerosis.* Neurol Sci, 2003. **24 Suppl 5**: p. S265-7.
18. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis.* Ann Neurol, 2001. **50**: p. 121-7.
19. Miller, D., et al., *Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis.* Lancet Neurol, 2005. **4**: p. 281-8.
20. Fernando, K.T., et al., *Magnetization transfer histograms in clinically isolated syndromes suggestive of multiple sclerosis.* Brain, 2005. **128**: p. 2911-25.
21. Prat, A., et al., *Pathogenesis of multiple sclerosis.* Curr Opin Neurol, 2005. **18**: p. 225-30.
22. Bjartmar, C., et al., *Axonal and neuronal degeneration in multiple sclerosis: mechanisms and functional consequences.* Curr Opin Neurol, 2001. **14**: p. 271-8.
23. Zamvil, S.S., et al., *Diverse targets for intervention during inflammatory and neurodegenerative phases of multiple sclerosis.* Neuron, 2003. **38**: p. 685-8.

24. Coles, A.J., et al., *Monoclonal antibody treatment exposes three mechanisms underlying the clinical course of multiple sclerosis.* Ann Neurol, 1999. **46**: p. 296-304.
25. Zeman, A.Z., et al., *A study of oligoclonal band negative multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1996. **60**: p. 27-30.
26. Aivado, M., et al., *Optimization and evaluation of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) with reversed-phase protein arrays for protein profiling.* Clin Chem Lab Med, 2005. **43**: p. 133-40.
27. You, J.S., et al., *The impact of blood contamination on the proteome of cerebrospinal fluid.* Proteomics, 2005. **5**: p. 290-6.
28. Ramstrom, M., et al., *Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry.* J Proteome Res, 2005. **4**: p. 410-6.
29. Dumont, D., et al., *Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients.* Proteomics, 2004. **4**: p. 2117-24.
30. Hammack, B.N., et al., *Proteomic analysis of multiple sclerosis cerebrospinal fluid.* Mult Scler, 2004. **10**: p. 245-60.
31. Kirkpatrick, L.L., et al., *Biochemical interactions of the neuronal pentraxins. Neuronal pentraxin (NP) receptor binds to taipoxin and taipoxin-associated calcium-binding protein 49 via NP1 and NP2.* J Biol Chem, 2000. **275**: p. 17786-92.
32. Stabach, P.R., et al., *Identification and characterization of beta V spectrin, a mammalian ortholog of Drosophila beta H spectrin.* J Biol Chem, 2000. **275**: p. 21385-95.
33. Reindl, M., et al., *Increased intrathecal production of apolipoprotein D in multiple sclerosis.* J Neuroimmunol, 2001. **119**: p. 327-32.
34. Mar, P., et al., *A longitudinal study of immunological parameters in multiple sclerosis: cell-mediated immunity and complement profiles.* Clin Exp Immunol, 1979. **36**: p. 442-8.
35. Williams, K.C., et al., *Immunology of multiple sclerosis.* Clin Neurosci, 1994. **2**: p. 229-45.
36. Li, W., et al., *Prenatal exposure of rats to Ginkgo biloba extract (EGb 761) increases neuronal survival/growth and alters gene expression in the developing fetal hippocampus.* Brain Res Dev Brain Res, 2003. **144**: p. 169-80.
37. Polihronis, M., et al., *Elevation of human cerebrospinal fluid clusterin concentration is associated with acute neuropathology.* J Neurol Sci, 1993. **115**: p. 230-3.
38. Zipp, F., et al., *The brain as a target of inflammation: common pathways link inflammatory and neurodegenerative diseases.* Trends Neurosci, 2006. **29**: p. 518-27.
39. Taylor, D.L., et al., *Activation of group II metabotropic glutamate receptors underlies microglial reactivity and neurotoxicity following stimulation with chromogranin A, a peptide up-regulated in Alzheimer's disease.* J Neurochem, 2002. **82**: p. 1179-91.



Chapter 5

Quantitative matrix-assisted laser desorption ionization fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple sclerosis-related proteins

Marcel P. Stoop, Lennard J. Dekker, Mark K. Titulaer, Robert-Jan A. N. Lamers, Peter C. Burgers, Peter A. E. Sillevis Smitt, Alain J. van Gool, Theo M. Luider, and Rogier Q. Hintzen

Journal of Proteome Research (2009), 8:1404-1414.

ABSTRACT

We introduce a MALDI-FT-ICR method for quantitative peptide profiling, using peak height as a measure for abundance. Relative standard deviations in peak height of peptides spiked over three orders of magnitude in concentration were below 10% and allowed for accurate comparisons between multiple sclerosis and controls. Application on a set of 163 CSF samples showed significantly differentially abundant peptides, which were subsequently identified into proteins (e.g. chromogranin A, clusterin, complement C3).

INTRODUCTION

The pathology of multiple sclerosis (MScl) is characterized by a combination of inflammation, demyelination and axonal damage [1]. Cerebrospinal fluid (CSF) is an promising body fluid in which to search for biomarkers and disease-associated proteins and peptides for MScl, due to its close proximity to disease processes [2, 3]. Additional advantages of CSF over serum or plasma are the overall lower protein concentration in CSF, which allows for easier observation of shedding or secretion of disease related proteins and the fact that CSF does not have an active clotting system as extensive as blood.

Recent CSF proteomics studies of MScl patients have identified proteins in pooled CSF using sample preparations steps such as ultraflitration [4] and 2D gel electrophoresis [5]. Other efforts have been made to identify biomarkers in subgroups of MScl [6], but later reports suggest that the proposed marker relates to a freezing artifact [7].

We have recently reported the results of CSF peptide profiling of MScl patients using MALDI-TOF mass spectrometry. With this technique we were able to discover specific peptide masses that were differentially abundant between MScl patients and non-inflammatory controls. In the present study we use advanced Fourier Transform mass spectrometry (FT-ICR MS) techniques to identify additional peptide masses, and to confirm by MALDI-FT-ICR MS our previous ELISA result that the protein chromogranin A was elevated in MScl patients [8]. Whereas the MALDI-TOF profiling study scored for presence of a peak in a number of replicate spectra, this MALDI-FT-ICR study records the height of the peaks for a quantitative analysis of differentially abundant peptides. A second aim of this study was to confirm the differentially abundant proteins found in our MALDI-TOF study by quantitative mass spectrometry.

MATERIALS AND METHODS

Patient selection

In brief, immediately after sampling, the CSF samples were centrifuged to discard cells and cellular elements (10 minutes at 3000 rpm). The samples were subsequently used for routine CSF diagnostics. Quantification of total protein was done by means of a turbidometric method on a Hitachi Modular P. Albumin and immunoglobulin G concentration were determined and the number of oligoclonal bands, which are present in most, but not all MScl patients, was recorded. Also quantification of the intrathecal cell count was reported. The remaining volume of

the samples was aliquoted and stored at -80°C, where they remained until sample preparation for this study. No extra freeze-thaw cycles were allowed. The CSF samples, which were all taken from untreated patients, were selected and classified into four groups by an experienced neurologist (RQH), based on all available clinical data. The groups were defined as follows; MScl: clinically defined multiple sclerosis according to the MacDonald criteria [9, 10] (n = 45), CIS: Clinically isolated syndrome of demyelination in one of four defined locations, or a combination thereof, which is in essence seen as a pre-stage of MScl [9, 10] (n = 40), OIND: inflammatory neurological diseases (n = 25), and OND: non-inflammatory neurological diseases (n = 53). Of the number of samples analyzed in this study 96% was previously used for peptide profiling by MALDI-TOF [8]. Six samples (1 Relapsing Remitting (RR) MScl, 1 CIS myelitis, and 4 controls (1 paraneoplastic syndrome, 1 infarct and 2 headaches)) were used to completion during the previous study and had to be replaced. Five new samples (2 RR MScl, 1 CIS optic neuritis/myelitis and two controls (1 Parkinson's disease and 1 headache)) were added to the sample set to keep the number of samples in each group roughly equal to the previous study. The influence of the new samples on the protein concentrations was neglectable. The Medical Ethical Committee of the Erasmus University Medical Center in Rotterdam, The Netherlands, approved the study protocol and all patients gave their consent.

Sample preparation

Prior to all sample preparation procedures, all samples were blinded. They were subsequently measured and analyzed in a random order, as described previously. In brief, twenty µl of each CSF sample was put into a 96-microtiter well plate (Nunc. Low binding, VWR, the Netherlands), and twenty µl 0.2% Rapigest (Waters, USA) in 50 mM ammonium bicarbonate buffer was added to each well. After 30 minute incubation periods with 1,4-dithiothreitol (60°C) and subsequently iodoacetamide (37°C), 4 µl 0.1 µg/µl gold grade trypsin (Promega, USA) / 3 mM Tris HCl (pH 8.0) was added to each well. The samples were incubated overnight at 37°C. To obtain a final concentration of 30-50 mM HCl (pH < 2), 2 µl of 500 mM HCl was added. The samples were then incubated for 45 minutes at 37°C, which stopped the digestion reaction.

Following the digestion procedure, the samples were desalted using 96-well zip C18 microtiter plates (Millipore, USA). The samples were centrifuged (Multifuge 3 S-R, Goffin Meyvis, the Netherlands) at 2000 rpm for five minutes. After the washing step, 3 µl acetonitrile were put on the C18 resin to prevent drying. Each trypsin digested CSF sample was mixed with 200 µl HPLC grade water / 0.1% TFA. The samples were subsequently put on to the washed and pre-wetted 96-well

zip C18 plate, and 30 minutes of centrifugation at 1500 rpm was used to load the peptides on to the C18 material. The wells were washed twice with 100 µl 0.1% TFA (5 minutes, 1750 rpm). An elution volume of 15 µl 50% acetonitrile/ 0.1% TFA was used to elute the samples in a new 96-well plate. The centrifugation period for the elution step was 30 minutes at 1600 rpm. After this elution step, the samples were stored in 96-well plates, which were covered with aluminum seals, at 4°C.

Quantitative measurement

A matrix solution was made by dissolving 10 mg 2,5-dihydroxybenzoic acid (DHB) in 1 ml 0.1% TFA/water. From each sample, 0.5 µl of elute was added to 0.5 µl of the matrix solution on the MALDI target (600/384 AnchorChip™ with transponder plate, Bruker Daltonics, Germany). All samples were spotted in duplicate. After drying at room temperature the samples were all measured manually on an APEX IV Qe 9.6 Tesla MALDI-FT-ICR mass spectrometer (Bruker Daltonics, USA) equipped with the first version of the vacuum Combisource, and a 20 Hz nitrogen laser with irradiation area of ~200 µm [11]. Xmass version 7.0.8 was used to operate the mass spectrometer, and DataAnalysis version 3.4 was used for data analysis (both from Bruker Daltonics). A multishot accumulation was used as recommended by Mize *et al.*, Moyer *et al.*, and O'Connor *et al.* [12-14]. Ten laser shots were accumulated in the storage hexapole, transferred to the FT-ICR cell, and scanned for 0.78 seconds. Fifty scans were summed for each mass spectrum. The acquisition mass range was m/z 800 to 4000. The mass spectra were subsequently apodized and zerofilled twice. An external mass calibration was applied using a quadratic equation.

All samples were measured until the highest peak in the spectrum (an albumin fragment at m/z 2045.0953) had attained a pre-determined intensity of 2*10e7. To determine the feasibility of using a quantitative approach to peak analysis a synthetic peptide (m/z 1616.7986, primary structure: Leu-Glu-Asn-Asn-Ala-Arg-Asp-Asp-Glu-Lys-Lys-Glu-Arg) was spiked into a CSF digest in 6 different concentrations. The height of the monoisotopic peak of the synthetic peptide was determined in all 6 mass spectra, which were measured in triplicate, and a calibration curve was made. Subsequently, a quality control sample was measured in triplicate to calculate a coefficient of variation and to assess the quality of the calibration curve.

Because variation in peak height in MALDI-FT-ICR mass spectrometry is much more reproducible as in MALDI-TOF mass spectrometry, we were able to use peak height as a means to quantitatively differentiate between the sample groups. Quantitative MALDI-FT-ICR has previously been applied to quantify HIV-1 protease inhibitors in cell lysates [15], indicating that quantitative MALDI-FT-ICR methods are readily applicable. The introduction of a quantitative peptide measurement method is a significant difference with the previously published study, which only scored

for presence or absence of peaks in a number of replicate spectra. The quantitative aspect of the results in the MALDI-FT-ICR study adds a new dimension to the peptide profiling results. Whereas previously the height of the peaks was not of influence on the statistical comparison between groups, the present study shows proteomics results of statistical comparison between groups based on quantitative measurements of peak height.

Analysis

The raw binary data files of the MALDI-FT-ICR experiments were loaded directly into the Eclipse software program (obtained from www.eclipse.org), which was adapted to compare mass spectra. All spectrum files were designated a group number (1 – MScl, 2 – CIS, 3 – OIND, 4 – OND) and serial numbered. Ten albumin peaks were used to calibrate the spectra. Of the 10 albumin peaks at least 8 had to be present in the spectrum, or else the spectrum would be excluded from the analysis. The algorithm described by Dekker and co-workers was used for peak detection, performed in the statistical language R (obtained from www.r-project.org), which has been included into the Eclipse software program [16-18]. The mass window (minimum distance between two adjacent peak positions) was set at 0.1 Da. After the generation of a peak list for each spectrum using the peak detection algorithm, an analyses matrix was created in R. During this process all peak lists of every sample were checked for the height of all peaks, which resulted in a matrix file that denoted a peak height for every peak position for every sample. A univariate analysis in R, in which a p-value was determined for every peak position, was used for statistical analysis, in which two groups were compared at a time, for a total of six individual comparisons between the groups. For data that does not follow a known theoretical distribution, one can use non-parametric tests. One such test is the Wilcoxon-Mann-Whitney test, which was used here for the pair-wise comparisons between groups. Data from multiple groups is combined and the data are tagged so that the group to which they belong is recorded. The combined data set is then sorted according to the value of the data. This type of statistical test is very well suited for the comparison of pre-defined groups, as was the case in these experiments [19, 20]. The peaks with low p-values ($p < 0.01$) in the comparisons were considered to be differentially abundant between the groups and were considered for identification purposes. Because of the non-parametric statistics used in the comparison between groups, the fold increase of every identified peptide with $p < 0.01$ in the comparisons was determined to confirm differential abundance between groups.

The statistical background of the comparisons was assessed by performing permutations. For each comparison 50 scramblings of the samples was performed

and thereafter the background and the related standard deviation of the background were determined. In these permutations the group numbers of the samples were scrambled prior to the Wilcoxon-Mann-Whitney test, which enabled us to define the number of background peptide peaks that was assigned a p-value below 0.01. By comparing this number with the number of peptide peaks that had low p-values in the real (not scrambled) comparisons, we were able to determine if there were statistically significant differences in peptide profiles between the groups. The cut-off number was set at the average value of background peaks ± 2 standard deviations (75 peptide peaks).

Identification

The next step in the process was identification of the differentially abundant peptides. Because of the very high number of peaks per spectrum, direct identification of the differentially abundant peptide peaks proved difficult in a MALDI-FT-ICR approach. For identification of the differentially abundant peptide peaks we used an online nano-LC pre-fractionation (C18 column) coupled with ESI-Orbitrap mass spectrometry. One microliter of the digested CSF was injected into the nano-LC (Ultimate3000 system, Dionex, USA) and the peptides were separated on a C18 PepMap 100 column (15 cm length, internal diameter 75 micrometer, Dionex, USA) after trapping the samples for 7.5 minutes on a C18 PepMap trapping column (5 cm length, internal diameter 300 micrometer), using a flow of 250 nl/min. Subsequently the peptides were measured in an Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Germany), using a data-dependent acquisition mode for the MS/MS identification step, which occurred in the ion trap. MS/MS was performed on the top 5 masses in the full scan spectra with a scan range from 200 to 2000 m/z. Active exclusion was used with a repeat count of 1, a repeat duration of 1 second, an exclusion list size of 500 masses and exclusion duration of 5 seconds. The MS/MS identifications were obtained using the Bioworks 3.2 software package (Thermo Fisher Scientific, Germany) and its SEQUEST feature, using the HUPO criteria, with XC scores of 1.8, 2.5 and 3.2 for single, double and triple charged ions respectively in the Swiss-Prot database (UniProtKB/Swiss-Prot Release 54.7).

We also used a nano-LC system (Ultimate3000 system, Dionex, USA) to separate the digested CSF peptides on a PepSwift Monolithic Column (200-µm x 50 mm, Dionex, USA), followed by off-line MALDI-TOF/TOF mass spectrometry. One microliter of digested CSF sample was injected and separated on this column during a 115 min LC run, using a 80 minute gradient of increasing concentrations of acetonitrile (Solvent A: 95% TFA (0.1%) in water, 5% ACN (v/v); Solvent B: 20% TFA (0.1%) in water, 80% ACN), and subsequently spotted on to a MALDI target plate. The MALDI target plate used in this nano-LC experiment was a pre-spotted

target plate (PAC 384 plate, Bruker Daltonics, Germany). On this plate the matrix was already applied during production and only the sample has to be added. Measurement of the pre-spotted plate was performed by automated MALDI-TOF/TOF (UltraFlex 3, Bruker Daltonics, Germany), using the Warp-LC software package, version 1.1 (Bruker Daltonics, Germany). Mass spectra for all spots were first generated and subsequently MS/MS spectra were generated in a data dependent manner. Only peaks that were suitable for MS/MS by a series of pre-set conditions, such as, for example, a sufficiently high signal-to-noise ratio (S/N > 5) were analyzed. By spotting DHB matrix solution on the PAC-plate and re-measuring the identical spots of interest by MALDI-FT-ICR, we confirmed the accurate masses of differentially abundant peptides [21]. Identifications were obtained by database searches of the MS/MS spectra using Mascot search program, (version 2.2) and the Swiss-Prot database (UniProtKB/Swiss-Prot Release 54.7), using the standard settings of the Warp-LC software package (IonScore > 15 for identifications). The mass accuracies (< 1 ppm) obtained in the MALDI-FT-ICR were used to further confirm the identification[22]. The mass accuracy of the masses observed in the MALDI-FT-ICR ought to be in a window less than 1 ppm of the calculated value of the peptide identified in the very spot that was investigated. For the Orbitrap a mass window of less than 2 ppm of the calculated value was used. Identification of peptides in this manner has recently been shown to have a 2.6% rate of false positive sequence matching [22], and that this is dependent on four parameters: the S/N ratio of the parent ion, the experimental mass accuracy, the statistical score of the identification, and the frequency of observation in the different samples measured

RESULTS

Clinical information

As indicated above, there was a 96% overlap between the samples used in this study and the samples used in the MALDI-TOF profiling study [8]. Details on male/female ratio of the samples, as well as information of age, protein concentrations (albumin, immunoglobulins and total) and diagnoses are listed in Table 1. The median time of disease duration for MScl patients at the time of CSF sampling was two years. The median elapsed time between the first occurrence of symptoms of the CIS patients and the time of CSF sampling was three months. The higher standard deviation of the CSF mean protein concentration of the OND group is due to a few samples, like for example two of the brain tumor samples, which have very high protein concentrations.

Table 1: Clinical information of all samples.

	MScl (n=45)	CIS (n=40)	OIND (n=25)	OND (n=53)
Male/Female ratio	12/33	12/28	7/19	23/31
% of samples positive for extra oligoclonal IgG bands	73.3%	57.5%	8.0%	3.8%
Mean Age[a]	41.7 (11.4)	33.9 (9.4)	49.1 (16.5)	48.1 (16.0)
Protein concentration (g/l)[a]	0.409 (0.140)	0.355 (0.107)	0.412 (0.161)	0.422 (0.217)
Albumin concentration (g/l)[a]	0.241 (0.100)	0.216 (0.079)	0.249 (0.160)	0.272 (0.191)
IgG concentration (g/l)[a]	0.061 (0.034)[b]	0.047 (0.030)[b]	0.055 (0.045)	0.043 (0.033)[b]
Diagnoses	- 31 RR MScl - 14 PP MScl	- 26 optic neuritis CIS - 6 myelitis CIS - 5 brainstem CIS - 3 other CIS	- 16 infections (bacterial and sterile) - 7 vasculitis - 2 Guillain Barré Syndrome	- 19 isolated headaches - 11 neurological degeneration - 10 nonneurological disease - 7 infarct - 3 brain tumor[c] - 3 other

a) Standard deviations in brackets.
b) p>0.05 for all comparisons (two-tailed t-test), except IgG concentrations for MScl-CIS (p=0.047) and MScl-OND (p=0.026).
c) The specific tumors are: 1 glioblastoma, 1 lumbar metastasis of an ependymoma and 1 meningioma.

Peak detection and data analysis

We determined if a linear correlation between the concentration of spiked peptides in CSF and peak height in MALDI-FT-ICR exists. Six concentrations of spiked peptide, over 3 orders of magnitude, were measured in triplicate. The measured peak heights of the six concentrations had an average CV of 5.9% (range 2.4-9.1%). From these six average peak heights that correlate with six concentrations of spiked synthetic peptide, we plotted a calibration curve. The correlation coefficient (R^2) of the calibration curve was 0.9955 (Figure 1). Measurement in triplicate of the quality control sample showed a CV of 10.8%, indicating that peak height does indeed correlate well to peptide concentration in a complex peptide mixture such as CSF. The sum of the height of 10 omnipresent albumin peaks correlated well with the albumin concentration measured by clinical chemistry ($R^2 = 0.9898$), again indicating a good correlation between peak height and peptide and protein concentration. Repeated measurement of an identical sample using a wide range of measurement times, based on the number of laser shots, showed that the background, which increases as the measurement times increase, did not influence the S/N ratio of the peptide peaks negatively. In fact, the S/N ratio of the peptide peaks increased when measurement times were prolonged. Due to the fact that background in MALDI-FT-ICR mass spectrometry using the vacuum Combisource, is only related to physical noise and not to chemical noise [15], there is also less

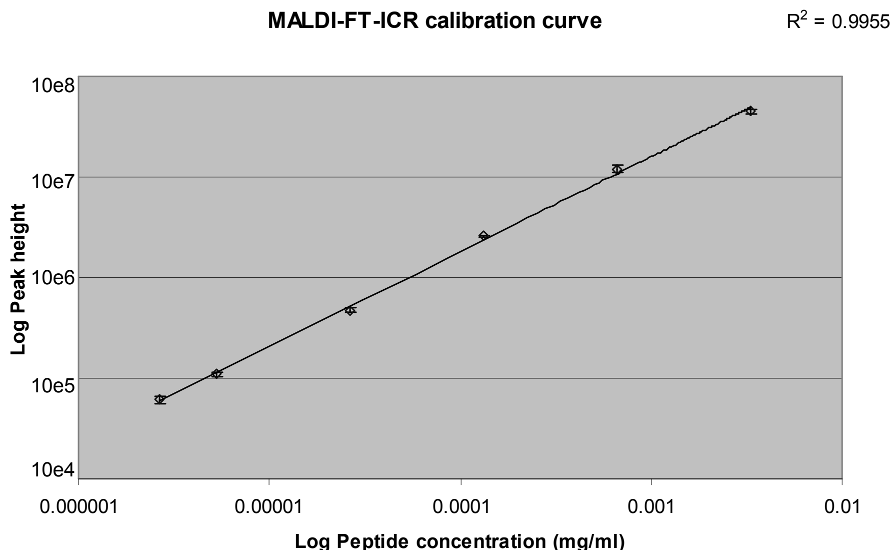
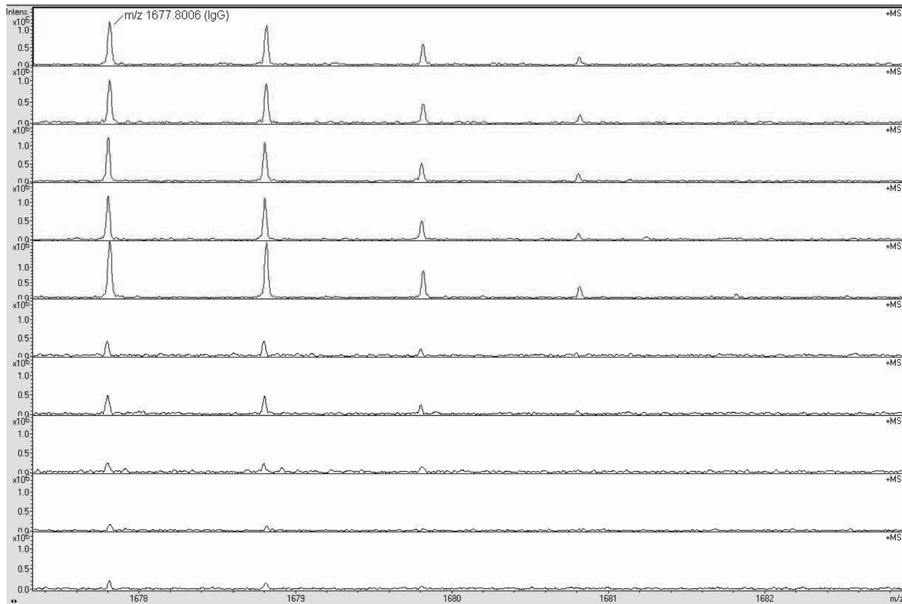


Figure 1. The calibration curve of the spiked CSF sample measured by MALDI-FT-ICR. The logarithmic scale is for visual purposes only. The correlation coefficient was obtained from a linear graph using a quadratic fit.

variation in both peak height and the height of noise peaks in MALDI-FT-ICR than in, for example, MALDI-TOF.

After the MALDI-FT-ICR spectra were loaded into Eclipse software package, they were each tagged with a group number (1 = MScl, 2 = CIS, 3 = OIND, 4 = OND). The spectra were calibrated using ten omnipresent albumin peaks and a peak matrix was generated in which the height of all peaks of every sample was recorded. The Wilcoxon-Mann-Whitney test was used to compare the groups and peaks with a low p-value ($p < 0.01$) were selected to be targeted for identification. In the comparison of MScl and OND samples 125 peptide peaks were present with $p < 0.01$. Among these masses that had a low p-value were three of the differentially abundant masses of the MALDI-TOF profiling study. These peaks, which correspond to peptides of clusterin, complement C3 and chromogranin A, were more often observed in the spectra of MScl patients than in OND control patients, in accordance with the results described previously [8]. The comparison of CIS and OND showed 93 peptide peaks with p-values below 0.01, while the comparisons of MScl and OIND, and CIS and OIND showed 73 and 107 peptide peaks respectively, with $p < 0.01$. The comparison of MS and CIS showed 83 peptide peaks with $p < 0.01$ and the comparison of OIND and OND showed 145 peptide peaks with low p-value. This resulted in a total of slightly over 600 differentially abundant peptide peaks, of which some overlap between multiple comparisons. Figure 2 shows a zoomed in sections of two differentially abundant peptides in the comparison of

2a



2b

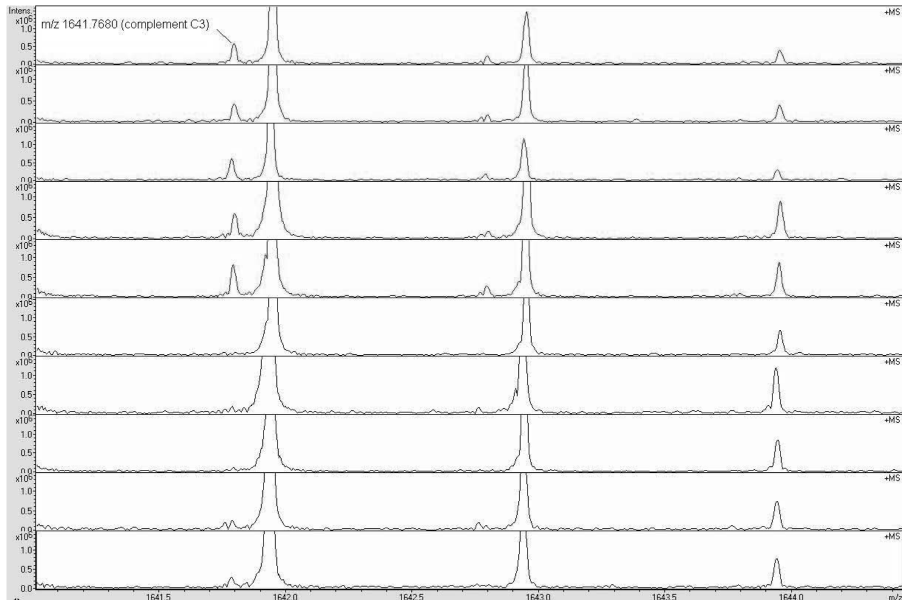


Figure 2. Zoomed in sections of five representative MScl (upper 5 spectra) and five representative OND (non-inflammatory control, lower 5 spectra) samples to highlight two differentially abundant peptides in the comparison of MScl and OND. Figure 2a shows a peptide peak at m/z 1677.8006, which was identified to be an Ig gamma-1 chain C region (P01857) peptide and Figure 2b shows a peptide peak at 1641.7680, which is directly adjacent to a large albumin peak at 1641.9377. The peptide peak at m/z 1641.7680 was identified to be a complement C3 (P01024) peptide.

MScl and OND (non inflamed controls). The background of this statistical procedure was determined by fifty randomizations of the samples. On average 61 (SD = 7) peaks had p-values lower than 0.01 when samples were randomly assigned to sample groups, whereas the number of peaks with $p < 0.01$ for the comparison between the pre-defined groups was between 73 and 145, depending on which groups are compared.

Identification

A total of 16 CSF samples were selected for nano-LC fractionation. These samples were selected randomly, except we matched for protein concentration. From all four groups samples from 2 males and 2 females were selected. These patients were diagnosed with a wide variety of diseases, such as, for example, 1 male and 1 female with RR MScl, 1 male and 1 female with PP MScl, a male with headaches, a male with a brain tumor, a female suffering from vasculitis and a female with a brain tumor. These samples were digested as described above and subsequently fractionated and measured by both the nano-LC-ESI-Orbitrap and the nano-LC-MALDI-TOF/TOF methods. One of the control sample experiments was not used due to a technical malfunction of the MS device, which left 15 samples for peptide identification purposes. The peptide identifications obtained were compared to the peak list of masses with low p-values to correlate these identifications with the differentially abundant peptides. An important observation which adds to the reliability of the CSF sampling and analysis is that known serum specific proteins such as hemoglobin and apolipoprotein B100, that ought to be absent in CSF, were not identified in these CSF samples, except in the brain tumor samples (hemoglobin). This indicates that the observed difference are indeed CSF specific and not due to either leakage or aberrant active transport from blood to CSF. This resulted in the identification of 89 peptide peaks with p-values below 0.01 on a total of approximately 600 peptide peaks with $p < 0.01$ (Table 2). Among these identifications were chromogranin A, clusterin, complement C3 and contactin-1, proteins previously reported in our MALDI-TOF study. Other identifications of differentially abundant peptide peaks included several specific regions of immunoglobulins as well as other proteins such as, for example, neuroendocrine protein 7B2 and pigment epithelium-derived factor. Sixteen identified peptides were found to be differentially abundant in the comparisons of multiple groups (Table 3).

Due to the quantitative nature of the MALDI-FT-ICR measurements it was possible to determine the increase in abundance of the differentially abundant peptides in the comparisons between groups. Most differentially abundant peptides ($p < 0.01$) were found to differ in intensity less than a factor 2 (median: 1.72, range: 1.27 - 4.43), but for 24 peptides the observed difference was larger

Table 2a: The identified peptides with p-values below 0.01 in the MScl-CIS comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1444.7382	4.1^{E-4}	MScl	2.70 (2.28-3.02)	P10909	Clusterin	RELDESLQVAER	0.55	MALDI-TOF/TOF
3267.5123	9.9^{E-3}	MScl	1.28 (1.08-1.41)	P41222	Prostaglandin-H2 D-isomerase	SPHWGSTYSVSV-VETDYDQYALLY-SQGSK	0.21	MALDI-TOF/TOF
2465.3786	5.7^{E-3}	CIS	1.55 (1.40-1.69)	P01019	Angiotensinogen	ADSQAQLLLSTV-VGVFTAPGLHLK	0.97	MALDI-TOF/TOF
1047.5937	7.5^{E-3}	CIS	1.41 (1.26-1.64)	Q12860	Contactin 1	HSIEVPIPR	0.76	MALDI-TOF/TOF
1226.6043	9.9^{E-3}	CIS	1.27 (1.14-1.39	P02768	Albumin	FKDLGEENFK	0.65	MALDI-TOF/TOF

a) p-values are denoted as 4.1^{E-3} instead of 4.1×10^{-3} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.

Table 2b. The identified peptides with p-values below 0.01 in the MScl-OIND comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1816.8907	1.3^{E-3}	MScl	1.91 (1.74-2.21)	P01024	Complement C3	SNLDEDIIAEENIVSR	0.88	MALDI-TOF/TOF
1909.9522	1.4^{E-3}	MScl	1.86 (1.63-2.02)	P41222	Prostaglandin-H2 D-isomerase	AQGFTEDTIV-FLPQTDK	0.99	MALDI-TOF/TOF
1632.7842	4.1^{E-3}	MScl	1.60 (1.45-1.69)	P18135	Ig kappa chain V-III region HAH	FSGSGSGTDFTLTISR	1.33[c]	MALDI-TOF/TOF, ESI-Orbitrap
1998.8325	2.1^{E-5}	OIND	3.22 (2.99-3.56)	P05408	Neuroendocrine protein 7B2	TADDGC#LENTPDT-AEFSR	1.00	MALDI-TOF/TOF
980.4942	3.5^{E-4}	OIND	2.01 (1.85-2.12)	P00738	Haptoglobin	VGYVSGWGR	0.62[c]	ESI-Orbitrap
1203.6362	1.7^{E-3}	OIND	1.97 (1.82-2.09)	P00738	Haptoglobin	VTSIQDWVQK	0.25	MALDI-TOF/TOF
1290.7286	1.7^{E-3}	OIND	1.95 (1.73-2.24)	P00738	Haptoglobin	DIAPTLTLYVGK	1.35[c]	MALDI-TOF/TOF, ESI-Orbitrap
2455.1491	5.0^{E-3}	OIND	1.67 (1.54-1.78)	P02766	Transthyretin	TSESGELHGLT-TEEEFVEGI YK	0.77	MALDI-TOF/TOF
2186.0389	8.4^{E-3}	OIND	1.33 (1.24-1.42)	P01009	Alpha-1-antitrypsin	LYHSEAFTVNFG-DTEEAKK	0.05	MALDI-TOF/TOF

a) p-values are denoted as 1.3^{E-3} instead of 1.3×10^{-3} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.
c) Mass accuracy was obtained by ESI-Orbitrap. All other mass accuracies were obtained by MALDI-FT-ICR.

Table 2c. The identified peptides with p-values below 0.01 in the MScl-OND comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1632.7848	2.2^{E-6}	MScl	4.43 (4.12-4.65)	P18135	Ig kappa chain V-III region HAH	FSGSGSGTDFTLTISR	1.33[c]	MALDI-TOF/TOF, ESI-Orbitrap
1677.8006	7.7^{E-5}	MScl	3.48 (3.19-3.68)	P01857	Ig gamma-1 chain C region	FNWYVDGVEVHNAK	0.41	MALDI-TOF/TOF
2074.9679	3.1^{E-4}	MScl	2.68 (2.50-2.83)	P10645	Chromogranin A	YPGPQAEGDSEGL-SQGLVDR	0.19	MALDI-TOF/TOF
1117.6089	3.2^{E-4}	MScl	2.62 (2.42-2.87)	P10909	Clusterin	TLLSNLEEAK	0.89	MALDI-TOF/TOF
1946.0291	1.3^{E-3}	MScl	1.98 (1.84-2.18)	Q6GMW0	IGKV1-5	TVAAPSVFIFPPSD-EQLK	1.09[c]	ESI-Orbitrap
1882.0047	1.4^{E-3}	MScl	1.96 (1.79-2.17)	P01766	Ig heavy chain V-III region BRO	EVQLVESGGGLVQP-GGSLR	1.01[c]	MALDI-TOF/TOF, ESI-Orbitrap
2139.0251	1.7^{E-3}	MScl	1.81 (1.68-1.95)	P01857	Ig gamma-1 chain C region	TPEVTC#-VVVDVSHEDPEVK	1.08[c]	MALDI-TOF/TOF, ESI-Orbitrap
1875.9251	2.1^{E-3}	MScl	1.80 (1.66-1.96)	P01834	Ig kappa chain C region	VYAC#-EVTHQGLSSPVTK	0.96	MALDI-TOF/TOF
1060.5420	2.1^{E-3}	MScl	1.79 (1.68-1.91)	P20774	Mimecan	DFADIPNLR	0.09	MALDI-TOF/TOF
1444.7382	2.3^{E-3}	MScl	1.77 (1.62-1.89)	P10909	Clusterin	RELDESLQVAER	0.55	MALDI-TOF/TOF
1743.8726	2.3^{E-3}	MScl	1.77 (1.61-1.88)	P41222	Prostaglandin-H2 D-isomerase	TMLLQPAGSLGSYSYR	0.70	MALDI-TOF/TOF
1909.9522	3.1^{E-3}	MScl	1.71 (1.50-1.89)	P41222	Prostaglandin-H2 D-isomerase	AQGFTEDTIVFLPQTDK	0.99	MALDI-TOF/TOF
1872.8486	3.4^{E-3}	MScl	1.68 (1.54-1.79)	P10909	Clusterin	QQTHMLDVMQDHFSR	0.37	MALDI-TOF/TOF
2651.2010	3.4^{E-3}	MScl	1.67 (1.52-1.81)	-	Conserved hypothetical protein	CYSCLVPAAHS-FVSSRSLSH NQEE	1.22[c]	ESI-Orbitrap
2564.2610	3.5^{E-3}	MScl	1.65 (1.51-1.79)	Q9UBP4	Dickkopf-related protein 3	GLLFPVC#-TPLPVEGELC#HDP ASR	0.98	MALDI-TOF/TOF
1641.7680	3.8^{E-3}	MScl	1.62 (1.48-1.75)	P01024	Complement C3	AGDFLEANYMNLQR	0.55	MALDI-TOF/TOF
973.5316	5.8^{E-3}	MScl	1.47 (1.38-1.56)	P01011	Alpha-1-antichymotrypsin	EQLSLLDR	0.31	MALDI-TOF/TOF
1647.7990	7.2^{E-3}	MScl	1.42 (1.31-1.52)	P02649	Apolipoprotein E	GEVQAMLGQSTEELR	0.97	MALDI-TOF/TOF
2157.0834	7.3^{E-3}	MScl	1.42 (1.30-1.51)	P01024	Complement C3	ILLQGTPVAQMTEDAV-DAER	1.02[c]	MALDI-TOF/TOF, ESI-Orbitrap
1148.6202	7.8^{E-3}	MScl	1.40 (1.26-1.49)	P01023	Alpha-2-macroglobulin	QGIPFFGQVR	0.78	MALDI-TOF/TOF
2439.2045	8.5^{E-3}	MScl	1.38 (1.23-1.49)	P18135	Ig kappa chain V-III region HAH	ASQSVSSSYLAWY-QQKPGQA PR	0.24	MALDI-TOF/TOF

Table 2c (continued)

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
3432.5922	8.6E-3	OND	1.37 (1.21-1.48)	P00738	Haptoglobin	AVGDKLPEC#EADDGC #PKPPEIAHGYVEHSVR	0.97	MALDI-TOF/ TOF
1883.9814	8.9E-3	OND	1.34 (1.19-1.46)	Q0IEY0	Tuberous sclerosis complex 2	TAVQQQQQQ-PTSQQR	1.99[c]	ESI-Orbitrap
2256.1624	9.2E-3	OND	1.33 (1.18-1.48)	P00450	Ceruloplasmin	KAEEEHLGILGPQL-HADVGD K	0.27	MALDI-TOF/ TOF

a) p-values are denoted as 2.2E-6 instead of 2.2×10^{-6} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.
c) Mass accuracy was obtained by ESI-Orbitrap. All other mass accuracies were obtained by MALDI-FT-ICR.

Table 2d: The identified peptides with p-values below 0.01 in the CIS-OIND comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
915.4562	1.5E-4	CIS	2.70 (2.48-2.96)	P02774	Vitamin D-binding protein	YTFELSR	1.31[c]	MALDI-TOF/ TOF, ESI-Orbitrap
2403.2069	9.4E-4	CIS	2.02 (1.84-2.23)	P25311	Zinc-alpha-2glycoprotein	HVEDVPAFQALG-SLNDLQFF R	0.91	MALDI-TOF/ TOF
2576.2409	2.9E-3	CIS	1.78 (1.63-1.95)	P47989	Xanthine dehydro-genase/oxidase	EAGEMELFVSTQNTM-KTQSFVAK	1.49[c]	ESI-Orbitrap
1816.8907	3.1E-3	CIS	1.74 (1.54-1.89)	P01024	Complement C3	SNLDEDIIAEENIVSR	0.88	MALDI-TOF/ TOF
1895.9916	4.0E-3	CIS	1.67 (1.54-1.87)	Q9BQT9	Calsyntenin-3	GHQPPPEMAGHS-LASSHR	0.63	MALDI-TOF/ TOF
1743.7811	6.4E-3	CIS	1.52 (1.39-1.66)	P41222	Prostaglandin-H2 D-isomerase	TMLLQPAGSLGSYSYR	0.70	MALDI-TOF/ TOF
2494.1475	7.5E-3	CIS	1.40 (1.28-1.53)	P01024	Complement C3	DYAGVFSDAGLTFTSSS-GQQ TAQR	0.20	MALDI-TOF/ TOF
1471.7426	9.9E-3	CIS	1.27 (1.15-1.35)	P01024	Complement C3	AAVYHHFISDGVR	0.95	MALDI-TOF/ TOF
1998.8325	3.1E-6	OIND	4.36 (4.03-4.49)	P05408	Neuroendocrine protein 7B2	TADDGC#-LENTPDTAEFSR	1.00	MALDI-TOF/ TOF
980.4948	7.1E-5	OIND	3.48 (3.31-3.69)	P00738	Haptoglobin	VGYVSGWGR	0.62[c]	ESI-Orbitrap
1290.7286	1.6E-4	OIND	2.85 (2.78-2.93)	P00738	Haptoglobin	DIAPTLTLYVGK	1.35[c]	MALDI-TOF/ TOF, ESI-Orbitrap
1203.6362	1.9E-3	OIND	2.12 (2.01-2.19)	P00738	Haptoglobin	VTSIQDWVQK	0.25	MALDI-TOF/ TOF
1716.7993	2.5E-3	OIND	1.81 (1.70-1.93)	Q12805	EGF-containing fibulin-like extracellular matrix protein 1	NPC#QDPYILTPENR	0.99	MALDI-TOF/ TOF

Table 2d (continued)

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1090.5679	4.9^{E-3}	OIND	1.62 (1.49-1.79)	P01009	Alpha-1-antitrypsin	WERPFEVK	0.09	MALDI-TOF/TOF
2291.1495	7.7^{E-3}	OIND	1.40 (1.29-1.48)	P01008	Antithrombin-III	AFLEVNEEGSEAAAS-TAVVI AGR	0.78	MALDI-TOF/TOF
1559.7854	8.0^{E-3}	OIND	1.35 (1.25-1.43)	P36955	Pigment epithelium-derived factor 1	LAAAVSNFGYDLYR	0.12	MALDI-TOF/TOF
2086.8354	9.4^{E-3}	OIND	1.30 (1.22-1.41)	P02768	Albumin	VHTEC#C#HGDLLEC#-ADDR	1.01[c]	MALDI-TOF/TOF, ESI-Orbitrap
1215.7316	9.5^{E-3}	OIND	1.28 (1.15-1.36)	P01011	Alpha-1-antichymotrypsin	ITLLSALVETR	0.74	MALDI-TOF/TOF
1803.9583	9.8^{E-3}	OIND	1.27 (1.12-1.43)	P01009	Alpha-1-antitrypsin	LQHLENELTHDIITK	0.89	MALDI-TOF/TOF, ESI-Orbitrap

a) p-values are denoted as 1.5^{E-4} instead of 1.5×10^{-4} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.
c) Mass accuracy was obtained by ESI-Orbitrap. All other mass accuracies were obtained by MALDI-FT-ICR.

Table 2e. The identified peptides with p-values below 0.01 in the CIS-OND comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1946.0291	9.1^{E-5}	CIS	3.27 (3.14-3.46)	Q6GMW0	IGKV1-5	TVAAPSVFIFPPSDE-QLK	1.09[c]	ESI-Orbitrap
1047.5937	1.4^{E-3}	CIS	2.18 (2.06-2.31)	Q12860	Contactin-1	HSIEVPIPR	0.76	MALDI-TOF/TOF
1677.8006	5.7^{E-3}	CIS	1.58 (1.44-1.69)	P01857	Ig gamma-1 chain C region	FNWYVDGVEVHNAK	0.36	MALDI-TOF/TOF
2651.2010	7.6^{E-3}	CIS	1.41 (1.29-1.56)	-	Conserved hypothetical prot.	CYSCLVPAAH-SFVSSRSLSH NQEE	1.22[c]	ESI-Orbitrap
1882.0047	8.5^{E-3}	CIS	1.32 (1.20-1.47)	P01766	Ig heavy chain V-III region BRO	EVQLVESGGG-LVQPGGSLR	1.01[c]	MALDI-TOF/TOF, ESI-Orbitrap
1883.9814	8.9^{E-3}	OND	1.30 (1.21-1.38)	Q0IEY0	Tuberous sclerosis complex 2	TAVQQQQQQ-PTSQQR	1.99[c]	ESI-Orbitrap

a) p-values are denoted as 9.1^{E-5} instead of 9.1×10^{-5} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.
c) Mass accuracy was obtained by ESI-Orbitrap. All other mass accuracies were obtained by MALDI-FT-ICR.

Table 2f. The identified peptides with p-values below 0.01 in the OIND-OND comparison.

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1998.8325	1.6^{E-5}	OIND	4.02 (3.82-4.23)	P05408	Neuroendocrine protein 7B2	TADDGC#LEN-TPDTAEFSR	1.00	MALDI-TOF/TOF
2047.8658	9.3^{E-5}	OIND	3.18 (2.98-4.35)	P02751	Fibronectin	GFNC#ESKPE-AEETC#FDK	0.73	MALDI-TOF/TOF
980.4942	1.8^{E-4}	OIND	2.86 (2.72-3.02)	P00738	Haptoglobin	VGYVSGWGR	0.62[c]	ESI-Orbitrap
1203.6365	2.0^{E-4}	OIND	2.80 (2.59-2.97)	P00738	Haptoglobin	VTSIQDWVQK	0.25	MALDI-TOF/TOF
1290.7286	3.4^{E-4}	OIND	2.70 (2.53-2.91)	P00738	Haptoglobin	DIAPTLTLYVGK	1.35[c]	MALDI-TOF/TOF, ESI-Orbitrap
1660.7650	3.5^{E-4}	OIND	2.67 (2.49-2.84)	P02765	Alpha-2-HS-glycoprotein	EHAVEGDC#D-FQLLK	0.90	MALDI-TOF/TOF
1955.9728	5.1^{E-4}	OIND	2.40 (2.23-2.58)	P36955	Pigment epithelium-derived factor 1	ALYYDLISSPDIHGT-YK	1.07[c]	MALDI-TOF/TOF, ESI-Orbitrap
1559.7854	6.5^{E-4}	OIND	2.27 (2.11-2.40)	P36955	Pigment epithelium-derived factor 1	LAAAVSNFGYDLYR	0.13	MALDI-TOF/TOF
1723.9621	9.4^{E-4}	OIND	2.01 (1.86-2.19)	P04217	Alpha-1B-glycoprotein	LELHVDGPPP-RPQLR	0.93	MALDI-TOF/TOF
1259.5718	1.5^{E-3}	OIND	1.91 (1.76-2.02)	P01023	Alpha-2-macroglobulin	VGFYESDVMGR	0.56	MALDI-TOF/TOF
2564.2610	1.5^{E-3}	OIND	1.92 (1.81-2.04)	Q9UBP4	Dickkopf-related protein 3	GLLFPVC#TPLPVEG-ELC#HDP ASR	0.97	MALDI-TOF/TOF
2086.8354	1.5^{E-3}	OIND	1.91 (1.74-2.06)	P02768	Albumin	VHTEC#C#HG-DLLEC#ADDR	1.01[c]	ESI-Orbitrap
1540.7289	1.6^{E-3}	OIND	1.88 (1.75-2.01)	P01876	Ig alpha-1 chain C region	DASGVTFTWT-PSSGK	0.72[c]	ESI-Orbitrap
1759.8681	1.7^{E-3}	OIND	1.86 (1.72-1.99)	P41222	Prostaglandin-H2 D-isomerase	TMLLQPAGSL-GSYSYR	0.23	MALDI-TOF/TOF
1215.7316	2.1^{E-3}	OIND	1.76 (1.62-1.89)	P01011	Alpha-1-antichymotrypsin	ITLLSALVETR	0.74	MALDI-TOF/TOF
1903.8200	2.8^{E-3}	OIND	1.64 (1.51-1.75)	P00450	Ceruloplasmin	NNEGTYYSPN-YNPQSR	0.26	MALDI-TOF/TOF
1946.0291	2.9^{E-3}	OIND	1.64 (1.48-1.73)	Q6GMW0	IGKV1-5	TVAAPSVFIFPPSDE-QLK	1.09[c]	ESI-Orbitrap
1914.8454	3.1^{E-3}	OIND	1.62 (1.48-1.70)	P00751	Complement factor B	FLC#TGGVSP-YADPNTC#R	0.99	MALDI-TOF/TOF
1921.0527	4.1^{E-3}	OIND	1.57 (1.46-1.68)	P01034	Cystatin C	KQIVAGVNYFLDVEL-GR	0.73	MALDI-TOF/TOF
1110.6044	5.4^{E-3}	OIND	1.48 (1.38-1.56)	P01009	Alpha-1-antitrypsin	LSITGTYDLK	0.25[c]	MALDI-TOF/TOF, ESI-Orbitrap
1148.5579	5.5^{E-3}	OIND	1.48 (1.37-1.56)	P61769	Beta-2-microglobulin	VEHSDLSFSK	0.26	MALDI-TOF/TOF

Table 2f (continued)

Peptide mass	p-value[a]	More abundant in	Fold change[b]	Accession code	Protein	Sequence	Mass accuracy (ppm)	Identification method
1117.6089	6.8^{E-3}	OIND	1.37 (1.29-1.49)	P10909	Clusterin	TLLSNLEEAK	0.89	MALDI-TOF/TOF
1716.7993	8.1^{E-3}	OIND	1.31 (1.24-1.40)	Q12805	EGF-containing fibulin-like extracellular matrix protein 1	NPC#QDPYILTPENR	0.99	MALDI-TOF/TOF
1220.6059	9.4^{E-3}	OIND	1.28 (1.19-1.37)	P02790	Hemopexin	NFPSPVDAAFR	0.05[c]	ESI-Orbitrap
1883.9814	8.0^{E-3}	OND	1.32 (1.23-1.41)	Q0IEY0	Tuberous sclerosis complex 2	TAVQQQQQQ-PTSQQR	1.99[c]	ESI-Orbitrap
1873.8530	9.5^{E-3}	OND	1.28 (1.21-1.36)	Q9P2N4	ADAMTS-9	IFGPGSQVCP-YMMQC#R	0.83[c]	ESI-Orbitrap

a) p-values are denoted as 1.6^{E-5} instead of 1.6×10^{-5} to limit the width of the table.
b) The fold change is the median increase (with the range in brackets) of peptide abundance in the group mentioned.
c) Mass accuracy was obtained by ESI-Orbitrap. All other mass accuracies were obtained by MALDI-FT-ICR.

than a two-fold increase. In the comparison between non-inflammatory controls and MScl the m/z 2074.9679 peptide of chromogranin A was found to be 2.69 times more abundant in MScl. A pair of immunoglobulin fragments was even more differentially abundant in this comparison, showing over three times higher abundance in CSF of MScl patients. Peptides with p-values higher than 0.01 (i.e. non-differentially abundant peptides) did not show increase in abundance. For example, the m/z 1067.6207 fragment of ubiquitin showed a p-value of 0.30 in the comparison between MScl and OND and the observed differential abundance was 1.17, and for instance the m/z 1620.8241 fragment of alpha-2-macroglobulin had a p-value of 0.61 in the same comparison and the observed differential abundance of this peptide was 1.09. In general peptide peaks with a p-value higher than 0.01 showed differential abundances in the range of 1.00 – 1.35 (median: 1.19). The differential abundances of peptide peaks with p-values smaller than 0.01 were all in the range of 1.27 – 4.43.

Discussion

Using peak height in MALDI-FT-ICR MS as a quantitative measurement for peptide abundance, it was possible to distinguish CSF of MScl (and CIS) patients from CSF of control patients, although the differences with non–inflammatory controls were more profound than the differences with inflammatory controls. When comparing both control groups to one another a large number of statistically significant differentially abundant peptide peaks is observed, which is, at least partially, due

Table 3. Peptides differentially abundant in more than one comparison (group with higher abundance in bold).

Comparison	Peptide mass	Accession code	protein
MScl-**OIND**, CIS-**OIND**, **OIND**-OND	980.4942	P00378	Haptoglobin
MScl-**OIND**, CIS-**OIND**, **OIND**-OND	1203.6362	P00378	Haptoglobin
MScl-**OIND**, CIS-**OIND**, **OIND**-OND	1290.7286	P00378	Haptoglobin
MScl-OIND, **CIS**-OIND	1816.8907	P01024	Complement C3
CIS-**OIND**, **OIND**-OND	2086.8354	P02768	Albumin
MScl-**OIND**, CIS-**OIND**, **OIND**-OND	1998.8325	P05408	Neuroendocrine protein 7B2
Mscl-OND, **OIND**-OND	1117.6089	P10909	Clusterin
MScl-CIS, **MScl**-OND	1444.7382	P10909	Clusterin
MScl-OIND, **MScl**-OND	1632.7842	P18135	Ig kappa chain V-III region HAH
CIS-**OIND**, **OIND**-OND	1559.7854	P36955	Pigment epithelium-derived factor 1
MScl-CIS, **MScl**-OIND	1909.9522	P41222	Prostaglandin-H2 D-isomerase
MScl-**OND**, CIS-**OND**, OIND-**OND**	1883.9814	Q0IEY0	Tuberous sclerosis complex 2
CIS-**OIND**, **OIND**-OND	1716.7993	Q12805	EGF-containing fibulin-like extracellular matrix protein 1
MScl-**CIS**, **CIS**-OND	1047.5937	Q12860	Contactin-1
MScl-OND, **OIND**-OND	1946.0291	Q6GMW0	IGKV1-5
MScl-OND, **OIND**-OND	2564.2610	Q9UBP4	Dickkopf-related protein 3

to inflammatory-related proteins found in the inflammatory control samples. The comparison of MScl and CIS showed few significantly differentially abundant peptides, which could fit with the majority of clinical observations that CIS is a pre-stage of MScl [9, 10]. It would be of interest to link MRI findings to these data. However, for this one would need standardised scanning protocols for assessing inflammation (e.g. gadolinium enhancement) and neurodegeneration. Such were not available in our material. Moreover to properly assess neurodegeneration, longitudinal MRI measurements would be needed, for example to quantify rate of atrophy and development of T1 hypo-intense lesions. Such studies are ongoing at a worldwide basis, and more intense combination with CSF sampling deserves to be recommended [23].

Protein differences in patient comparison studies are virtually never black-and-white phenomena. Thus compounds will mostly still be present in a non-diseased state, but in changed concentrations [24-27]. While scoring presence and non-presence of peptide peaks certainly is a viable method of monitoring differential expression [17], it seems more elegant and more elucidating to monitor differential peptide expression in a quantitative way by measuring peak height. Consequently, methods for monitoring peak height must be able to relate peak height to peptide concentration in a complex sample. While another report has shown the applicability of MALDI-FT-ICR mass spectrometry as a quantitative proteomics tool [28], a quantitation test was performed on peptide mixtures in CSF. Using an APEX IV Qe

9.6 Tesla MALDI-FT-ICR mass spectrometer we were able to measure the intensity of a synthetic peptide, which was spiked into a CSF sample in 6 concentrations, over 3 orders of magnitude. The positive correlation between relative concentration and intensity had an R^2 of 0.9955, and a spiked quality control sample showed a CV of 10.8%. This rather good correlation between peak height and concentration in MALDI-FT-ICR, allowed us to relate peak height to concentration in a meaningful way, as opposed to, for example MALDI-TOF, which shows far larger CV's in complex peptide mixtures [17] and which consequently is only yet applicable as a semi-quantitative method.

Several other research groups have previously reported on the CSF proteomic profile of MScl. Some studies were set up to identify proteins regardless of difference with proteomic profiles of controls [29, 30]. Other studies compared CSF of MScl patients to a homogeneous group of patients afflicted with another disease [31], whereas yet another group performed differential proteomics using 2D gel electrophoresis on a smaller number of patients [32]. The latter study is especially interesting for comparison with our data, as the objective of this study was similar to this study. Although most differentially abundant proteins reported by Lehmensiek *et al.* [32] were also identified in our study, the majority of them were not statistically significant in our sample groups examined. However, the presence of elevated levels of Ig kappa chain and complement component C3 in both sample sets was confirmed.Of the eleven identified peptides found by the MALDI-TOF profiling study [8], seven were again identified using the MALDI-FT-ICR mass spectrometer, three of those with p-values below 0.01 (Table 4). The peptides of chromogranin A, clusterin and complement C3 were again found with low p-values, confirming the previous observations. The other four identified peptides were found with non-significant p-values, indicating that they were not found to be differentially abundant using peak intensity. Another peptide of complement C4A was found to be differentially abundant in the comparison of CIS with inflammatory controls. The differences between the two studies may be due to several factors. First, the previous analysis scored for presence and absence of peaks, using 6 replicas of each sample. The method of analysis described here takes into account the peak height. Second, since orthogonal MALDI mass spectrometry

Table 4. Overlap of identified peptides with the identifications previously reported in the MALDI-TOF profiling study. All three peptides were more abundant in MScl, in both studies.

Comparison	Protein	Accession code	Peptide sequence	Measured mass	Mass accuracy (ppm)
MScl-OND	Chromogranin A	P10645	YPGPQAEGDSEGLSQGLVDR	2074.9679	0.19
MScl-OND	Clusterin	P10909	TLLSNLEEAK	1117.6089	0.89
MScl-OND	Complement C3	P01024	AGDFLEANYMNLQR	1641.7680	0.55

is not compatible with α-cyano-4-hydroxycinnamic acid as a matrix, a different matrix was used in these experiments, namely DHB. It is well known that different matrices may lead to very different ionization efficiencies [33]. In fact, we analyze a certain window of proteins related to a matrix and not all the proteins that might be expected in CSF.

Because of ethical issues regarding CSF sampling it is hard to obtain a sufficient number of CSF samples from healthy patients. The samples in the control groups used in this study were taken from patients with a wide variety of diagnoses other than MScl, or its pre-stage, CIS. This could add to the robustness of the study, since the analysis still picks out specific peptides that are significant differentially abundant in MScl compared to broad ranged control groups. It stands to reason that some control samples influence the comparison more than others. One of the peptides that was found to be more abundant in OND than in MScl is tuberous sclerosis complex 2, a protein known to be aberrantly abundant in brain tumors [34]. There are three brain tumors present in the non-inflammatory control group, two of which (a glioblastoma and a lumbar metastasis of an ependymoma) have very high peaks at the mass that relates to the identified peptide of tuberous sclerosis complex 2, while other control samples do not have high peptide peaks at this particular mass.

The ESI-Orbitrap is a tool that is well suited for peptide and protein identification and recent reports also show that it is suited for peptide profiling [35]. However, measuring and analyzing 163 samples on the ESI-Orbitrap is very time consuming, while measuring the same amount of samples on the MALDI-FT-ICR is quite fast, which is highly preferable in peptide profiling studies due to the number of samples that needs to be measured for meaningful statistical comparison between sample groups

The average background number of peptide peaks was obtained by randomisation of the sample group numbers prior to the Wilcoxon-Mann-Whitney test. There were 61 ± 7 background peaks with a p-value below 0.01. This indicates that a number of identified, differentially abundant peptides could be due to statistical background. When comparing the number of differentially abundant peptide peaks found in the comparisons to the number of background peaks, it seems that some comparisons, such as MScl versus non-inflammatory controls, showed statistically more differences (125) than other comparisons, like MScl versus inflammatory controls (107). However, due to the fact that MScl is believed to have an active inflammatory component [1], this observation can be expected.

While most of the identified differentially abundant proteins were found either by a single peptide with a p-value below 0.01 or by a small number of peptides with a p-value below 0.01, the identification of the albumin peak at m/z 2086.8354

stands out. However, only 5.3% (2/38) of the identified albumin peaks have significant p-values, all other albumin identifications related to peptide masses present in the FT-ICR profiling (36/38) were not differentially expressed. Therefore we cannot conclude albumin to be significantly more abundant in one of the groups in these comparisons. Also, it must be noted that the eight patients of the OIND group that have the highest peak at m/z 2086.8354 are also the samples with the highest albumin concentration in the non-inflammatory control group. In contrast to the 2/38 albumin peptides found to be differentially abundant, four different peptide masses of haptoglobin were identified in de peptide profiling matrix. In the comparison between MScl and the inflammatory controls three of these four peptides were differentially abundant ($p < 0.01$).

The implication of immunoglobulins in MScl pathogenesis has been reported repeatedly. Elevated immunoglobulin G (IgG) levels as a marker for disease activity [36, 37], extra oligoclonal IgG bands [38] as a potential marker for progression from CIS to MScl [39] and free immunoglobulin kappa light chains in MScl cerebrospinal fluid [40-42] have all been extensively researched. Considering the fact that immunoglobulin levels in the MScl samples used in this study were significantly higher than in non-inflammatory controls, it was expected that some immunoglobulin peptides would be identified as differentially abundant using this MALDI-FT-ICR analysis method. The identification of peptides of two immunoglobulin kappa light chains with elevated expression in MScl CSF is also interesting in view of previous suggestions that these molecules may serve as a marker for progression [40]. We have now identified the amino acid sequences of this marker. It appears that they are derived from the invariable region of the immunoglobulin molecule. Although concentrations are expected to be very low, mass spectrometry has been shown to be effective in sequencing immunoglobulin regions [43]. We only identified 14.5% of the differentially abundant peptide peaks in this study. So these variable regions of immunoglobulins will have to be identified among the remaining 85.5%.

The finding of elevated expression of haptoglobin peptides in the inflammatory diseases is in line with previously published reports on acute phase inflammation markers [44, 45]. The detection of the previously reported three peptides of chromogranin A, clusterin en complement C3 (Table 4) shows that these peptides are differentially abundant in these samples regardless of the mass spectrometry method used for detection. Clusterin and complement C3 have been repeatedly reported in MScl studies as being elevated in body fluids, and both proteins are functional in inflammatory processes, which are considered pivotal for the pathogenesis of MScl [46-49]. A recent study reported elevated chromogranin B levels in CSF of MScl patients [26], whereas we reported previously that chromogranin A was present in elevated levels in CSF of MScl patients [8], and a recent re-

port showed chromogranin A to be present in MScl lesions [50], so this family of secretogranins could be an interesting protein family to investigate in MScl pathogenesis. Chromogranin A levels have previously been reported to differ in CSF of patients with Parkinson's disease [51]. One sample of a patient with Parkinson's disease was included in the non-inflammatory control group. This sample also showed an intense peptide peak at m/z 2074.9679, the mass of the identified chromogranin A peptide. It has been reported that chromogranin A can induce neurotoxicity in migroglial cells [52], so it would be of interest to see if higher chromogranin A levels are associated with more aggressive clinical disease types. This molecule may very well be an important component of the innate immune arm that is increasingly implicated in neurodegenerative processes [53].

ACKNOWLEDGEMENTS

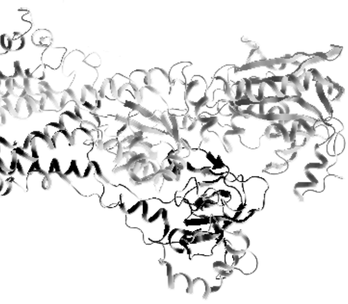
This work was supported by a program grant provided by the Dutch Multiple Sclerosis Research Foundation to the Rotterdam Multiple Sclerosis research Center, ErasMS, the Netherlands Organization for Scientific research (ZON-MW, RQH), the Netherlands Proteomics Centre and the Top Institute Pharma grant D4-102-1.

REFERENCES

1. Compston, A., et al., *Multiple sclerosis.* Lancet, 2002. **359**: p. 1221-31.
2. Romeo, M.J., et al., *CSF proteome: a protein repository for potential biomarker identification.* Expert Rev Proteomics, 2005. **2**: p. 57-70.
3. Yuan, X., et al., *Proteomics analysis of prefractionated human lumbar cerebrospinal fluid.* Proteomics, 2005. **5**: p. 541-50.
4. Noben, J.P., et al., *Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography, and mass spectrometry.* J Proteome Res, 2006. **5**: p. 1647-57.
5. Hammack, B.N., et al., *Improved resolution of human cerebrospinal fluid proteins on two-dimensional gels.* Mult Scler, 2003. **9**: p. 472-5.
6. Irani, D.N., et al., *Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis.* Ann Neurol, 2006. **59**: p. 237-47.
7. Del Boccio, P., et al., *Cleavage of cystatin C is not associated with multiple sclerosis.* Ann Neurol, 2007. **62**: p. 201-4; discussion 205.
8. Stoop, M.P., et al., *Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry.* Proteomics, 2008. **8**: p. 1576-85.
9. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis.* Ann Neurol, 2001. **50**: p. 121-7.
10. Miller, D., et al., *Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis.* Lancet Neurol, 2005. **4**: p. 281-8.
11. Baykut, G., et al., *Matrix-assisted laser desorption/ionization fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation.* Rapid Commun Mass Spectrom, 2000. **14**: p. 1238-47.
12. Mize, T.H., et al., *Broad-band ion accumulation with an internal source MALDI-FTICR-MS.* Anal Chem, 2000. **72**: p. 5886-91.
13. Moyer, S.C., et al., *Attomole peptide analysis by high-pressure matrix-assisted laser desorption/ionization Fourier transform mass spectrometry.* Anal Chem, 2003. **75**: p. 6449-54.
14. O'Connor, P.B., et al., *Application of multishot acquisition in Fourier transform mass spectrometry.* Anal Chem, 2000. **72**: p. 5125-30.
15. van Kampen, J.J., et al., *Quantitative analysis of HIV-1 protease inhibitors in cell lysates using MALDI-FTICR mass spectrometry.* Anal Chem, 2008. **80**: p. 3751-6.
16. Dekker, L.J., et al., *MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancer.* Mol Cell Proteomics, 2005. **4**: p. 1341-9.
17. Dekker, L.J., et al., *A new method to analyze matrix-assisted laser desorption/ionization time-of-flight peptide profiling mass spectra.* Rapid Commun Mass Spectrom, 2005. **19**: p. 865-70.
18. Titulaer, M.K., et al., *A software application for comparing large numbers of high resolution MALDI-FTICR MS spectra demonstrated by searching candidate biomarkers for glioma blood vessel formation.* BMC Bioinformatics, 2008. **9**: p. 133.
19. Gaddis, G.M., et al., *Introduction to biostatistics: Part 5, Statistical inference techniques for hypothesis testing with nonparametric data.* Ann Emerg Med, 1990. **19**: p. 1054-9.

20. Titulaer, M.K., et al., *A database application for pre-processing, storage and comparison of mass spectra derived from patients and controls.* BMC Bioinformatics, 2006. **7**: p. 403.
21. Dekker, L.J., et al., *FTMS and TOF/TOF mass spectrometry in concert: identifying peptides with high reliability using matrix prespotted MALDI target plates.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **847**: p. 62-4.
22. Stoop, M.P., et al., *The rate of false positive sequence matches of peptides profiled by MALDI MS and identified by MS/MS.* Journal of Proteome research (In Press), 2008.
23. Nielsen, J.M., et al., *Diagnosing MS: recent guidelines and future goals focusing on magnetic resonance imaging.* Int MS J, 2007. **14**: p. 29-34.
24. Armah, H.B., et al., *Cerebrospinal fluid and serum biomarkers of cerebral malaria mortality in Ghanaian children.* Malar J, 2007. **6**: p. 147.
25. Kalinowska, A., et al., *PECAM-1, a key player in neuroinflammation.* Eur J Neurol, 2006. **13**: p. 1284-90.
26. O'Connor, K.C., et al., *Comprehensive phenotyping in multiple sclerosis: discovery based proteomics and the current understanding of putative biomarkers.* Dis Markers, 2006. **22**: p. 213-25.
27. Petzold, A., et al., *A systematic review and meta-analysis of CSF neurofilament protein levels as biomarkers in dementia.* Neurodegener Dis, 2007. **4**: p. 185-94.
28. Ma, M., et al., *Methyl esterification assisted MALDI FTMS characterization of the orcokinin neuropeptide family.* Anal Chem, 2007. **79**: p. 673-81.
29. Dumont, D., et al., *Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients*. Proteomics, 2004. **4**: p. 2117-24.
30. Hammack, B.N., et al., *Proteomic analysis of multiple sclerosis cerebrospinal fluid.* Mult Scler, 2004. **10**: p. 245-60.
31. D'Aguanno, S., et al., *Differential cerebro spinal fluid proteome investigation of Leber hereditary optic neuropathy (LHON) and multiple sclerosis.* J Neuroimmunol, 2008. **193**: p. 156-60.
32. Lehmensiek, V., et al., *Cerebrospinal fluid proteome profile in multiple sclerosis.* Mult Scler, 2007. **13**: p. 840-9.
33. Keller, B.O., et al., *Three-layer matrix/sample preparation method for MALDI MS analysis of low nanomolar protein samples.* J Am Soc Mass Spectrom, 2006. **17**: p. 780-5.
34. Farrell, C.J., et al., *Genetic causes of brain tumors: neurofibromatosis, tuberous sclerosis, von Hippel-Lindau, and other syndromes.* Neurol Clin, 2007. **25**: p. 925-46, viii.
35. Scherl, A., et al., *Genome-Specific Gas-Phase Fractionation Strategy for Improved Shotgun Proteomic Profiling of Proteotypic Peptides.* Anal Chem, 2008.
36. Caroscio, J.T., et al., *Quantitative cerebrospinal fluid IgG measurements as a marker of disease activity in multiple sclerosis.* Arch Neurol, 1986. **43**: p. 1129-31.
37. Mehta, P.D., et al., *Quantitation of IgG and albumin in CSF and serum from multiple sclerosis patients by enzyme-linked immunosorbent assay.* Acta Neurol Scand, 1987. **75**: p. 125-9.
38. Zeman, A.Z., et al., *A study of oligoclonal band negative multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1996. **60**: p. 27-30.
39. Lim, E.T., et al., *Cerebrospinal fluid levels of brain specific proteins in optic neuritis.* Mult Scler, 2004. **10**: p. 261-5.
40. Rinker, J.R., 2nd, et al., *Elevated CSF free kappa light chains correlate with disability prognosis in multiple sclerosis.* Neurology, 2006. **67**: p. 1288-90.

41. Fagnart, O.C., et al., *Free kappa and lambda light chain levels in the cerebrospinal fluid of patients with multiple sclerosis and other neurological diseases.* J Neuroimmunol, 1988. **19**: p. 119-32.
42. Rudick, R.A., et al., *Free kappa light chains in multiple sclerosis spinal fluid.* Ann Neurol, 1986. **20**: p. 63-9.
43. Obermeier, B., et al., *Matching of oligoclonal immunoglobulin transcriptomes and proteomes of cerebrospinal fluid in multiple sclerosis*. Nat Med, 2008. **14**: p. 688-93.
44. Giffen, P.S., et al., *Markers of experimental acute inflammation in the Wistar Han rat with particular reference to haptoglobin and C-reactive protein.* Arch Toxicol, 2003. **77**: p. 392-402.
45. Steel, D.M., et al., *The major acute phase reactants: C-reactive protein, serum amyloid P component and serum amyloid A protein.* Immunol Today, 1994. **15**: p. 81-8.
46. Li, W., et al., *Prenatal exposure of rats to Ginkgo biloba extract (EGb 761) increases neuronal survival/growth and alters gene expression in the developing fetal hippocampus.* Brain Res Dev Brain Res, 2003. **144**: p. 169-80.
47. Mar, P., et al., *A longitudinal study of immunological parameters in multiple sclerosis: cell-mediated immunity and complement profiles.* Clin Exp Immunol, 1979. **36**: p. 442-8.
48. Polihronis, M., et al., *Elevation of human cerebrospinal fluid clusterin concentration is associated with acute neuropathology.* J Neurol Sci, 1993. **115**: p. 230-3.
49. Williams, K.C., et al., *Immunology of multiple sclerosis.* Clin Neurosci, 1994. **2**: p. 229-45.
50. Han, M.H., Hwang S., Roy, D. B., Lundgren, D. H., *et al.*, *Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets.* Nature, 2008. **451**: p. 1076-1081.
51. Eder, U., et al., *Levels and proteolytic processing of chromogranin A and B and secretogranin II in cerebrospinal fluid in neurological diseases.* J Neural Transm, 1998. **105**: p. 39-51.
52. Taylor, D.L., et al., *Activation of group II metabotropic glutamate receptors underlies microglial reactivity and neurotoxicity following stimulation with chromogranin A, a peptide up-regulated in Alzheimer's disease.* J Neurochem, 2002. **82**: p. 1179-91.
53. Infante-Duarte, C., et al., *New developments in understanding and treating neuroinflammation.* J Mol Med, 2008. **86**: p. 975-85.



Chapter 6

Proteomics comparison of CSF in Relapsing Remitting and Primary Progressive Multiple Sclerosis

Marcel P. Stoop, Vaibhav Singh, Lennard J. Dekker, Mark K. Titulaer, Christoph Stingl, Peter C. Burgers, Peter A. E. Sillevis Smitt, Rogier Q Hintzen, and Theo M. Luider.

Manuscript submitted for publication.

ABSTRACT

We examined the CSF proteomic profile of two different multiple sclerosis (MScl) subgroups; relapsing remitting (RR) MScl and primary progressive (PP) MScl, with the aim to find proteins that differentiate between the two disease types. We analyzed 10 PP MScl, 11 RR MScl and 10 neurologically and internally healthy non-neurological control CSF samples by quantitative MALDI-FT-ICR mass spectrometry after tryptically digesting the proteins in the CSF into peptides. Using the Wilcoxon-Mann-Whitney test to compare the groups pair-wise, the comparison between both MScl types resulted in 15 peak masses with p-values below 0.01, of which 7 were identified by ESI-Orbitrap mass spectrometry. Two of these proteins were validated by immunoassay. Protein jagged-1 (Western blot, p = 0.0124) and vitamin D-binding protein (ELISA, p= 0.0058) were validated to be significantly more abundant in RR multiple sclerosis than in PP multiple sclerosis. The jagged-1 protein has previously been implicated in multiple sclerosis because of its role in neuronal homeostasis and in CD4+ as well as CD8+ T-cell activation and differentiation. Vitamin D-binding protein has, next to its function in vitamin D homeostasis, other activities with possible relevance for MScl pathology, such as complement-guided chemotaxis and stimulatory effects on macrophages.

INTRODUCTION

Multiple sclerosis (MScl) can be divided into two major subtypes based on clinical representation of disease symptoms in the patients [1]. Between 85-90% of patients can be classified as having the relapsing remitting (RR) MScl subtype, in which disease relapses are followed by periods of remission, and 10-15% of all MScl patients are diagnosed with the primary progressive (PP) subtype [2]. Even within a single large Dutch MScl pedigree of 26 patients with similar genetic background, the percentage of patients with a PP phenotype remained 15% [3].

By definition, in PP patients disease progression is characterized by a progressive course without relapses or remissions from the onset of the disease [4]. PP patients tend to have lower inflammatory lesional activity, for which no immunological or genetic explanation has been identified yet. The scarce comparative neuropathological studies show a large overlap in lesional pathology, but indicate less inflammatory activity for PP, with still substantial axonal damage [5]. The general picture is that relapse onset and PP forms share substantial characteristics. In other words, it has remained a challenge to identify the biological parameters that determine a PP disease course.

Although proteomic analysis of active multiple sclerosis lesions may be a straightforward approach to study the processes involved in MScl disease pathways [6], this is very difficult to perform in living individuals. In most cases the pathology of the disease can only be investigated in post-mortem material, which quite frequently represents the end-stage of the disease. The study of CSF taken during disease appears a good alternative. CSF is in close contact with the CNS parenchyma and collects the products of the inflammatory and neurodegenerative processes of MScl activity.

Proteomics analysis of CSF has detected a number of proteins that were elevated in MScl patients. Additionally, differentially abundant proteins identified by proteomics, such as apolipoprotein A1 [7] and chromogranin A [8] were validated by other techniques. Other studies added additional data on elevated immunoglobulin expression in MScl CSF, as well as increased levels of apolipoprotein E [9, 10]. Yet in all currently reported proteomics CSF studies of MScl patients either only a single subtype of MScl patients or a combined group of all subtypes of MScl was studied.In the current study we specifically aimed to differentiate between the MScl patients and completely healthy controls and between both subtypes of MScl by comparing CSF proteins.

MATERIALS AND METHODS

Patient selection

The CSF samples of MScl patients, divided into two groups, RR MScl and PP MScl, were collected from untreated patients undergoing routine diagnostic procedures by an experienced neurologist (RQH). The healthy control CSF samples were taken from patients receiving spinal anesthesia prior to non-neurological minor surgical interventions, such as knee and hip replacements, groin rupture and Achilles tendon rupture.

All samples were handled in exactly the same way after sampling, using a procedure that has been previously reported [10]. In brief, immediately after sampling, the CSF samples were centrifuged to discard cells and cellular elements and the total protein concentration and albumin concentrations were determined. The number of oligoclonal bands and the intrathecal cell count were also reported. The remaining volume of the samples was aliquoted and stored at -80ºC, where they remained until sample preparation for this study. No extra freeze-thaw cycles were allowed. The Medical Ethical Committee of the Erasmus University Medical Centre in Rotterdam, The Netherlands, approved the study protocol and all patients gave their consent.

Sample preparation, measurement and analysis

The CSF samples were handled according to the same protocol for quantitative MALDI-FT-ICR MS measurements we reported previously [10], which consists of a blinded experiment in which the samples were digested by trypsin and subsequently measured by MALDI-FT-ICR. After calibration by means of omnipresent albumin peaks, an analysis matrix is generated. A univariate analysis, in which a p-value was determined for every peak position, was used for statistical analysis, in which two groups were compared at a time, for a total of three individual comparisons between the groups. The differentially abundant peaks ($p < 0.01$) in the comparisons were considered for identification purposes. The fold increase of every identified peptide with $p < 0.01$ in the comparisons was determined to confirm differential abundance between groups. Assessment of the statistical background, by means of permutation of a series of 50 scramblings of the samples for each comparison, was used to define a cut-off number for the determination of statistically significant differences between groups.

The differentially abundant peptides were identified by nano-LC-ESI-Orbitrap MS. These measurements were carried out on a Ultimate 3000 nano LC system (Dionex, Germering, Germany) online coupled to a hybrid linear ion trap / Orbitrap MS (LTQ Orbitrap XL; Thermo Fisher Scientific, Germany). Five µL digest were

loaded on to a C18 trap column (C18 PepMap, 300µm ID x 5mm, 5µm particle size, 100 Å pore size; Dionex, The Netherlands) and desalted for 10 minutes using a flow rate of 20 µL /min 0.1% TFA. Then the trap column was switched online with the analytical column (PepMap C18, 75 µm ID x 150 mm, 3 µm particle and 100 Å pore size; Dionex, The Netherlands) and peptides were eluted with following binary gradient: 0% - 25% solvent B in 120 min and 25% - 50% solvent B in further 60 minutes, where solvent A consist of 2% acetonitrile and 0.1% formic in water and solvent B consists of 80% acetonitrile and 0.08% formic acid in water. Column flow rate was set to 300 nL/min. For MS detection a data dependent acquisition method was used: high resolution survey scan from 400 – 1800 Th. was performed in the Orbitrap (value of target of automatic gain control AGC 10^6, resolution 30,000 at 400 m/z; lock mass was set to 445.120025 u (protonated $(Si(CH_3)_2O)_6$) [11]). Based on this survey scan the 5 most intensive ions were consecutively isolated (AGC target set to 10^4 ions) and fragmented by collision-activated dissociation (CAD) applying 35% normalized collision energy in the linear ion trap. After precursors were selected for MS/MS, they were excluded for further MS/MS spectra for 3 minutes. The MS/MS identifications were obtained using in the Bioworks 3.2 (peak picking by Extract_msn, default settings) software package (Thermo Fisher Scientific, Germany), and its SEQUEST feature, using the HUPO criteria, with XC scores of 1.8, 2.2 and 3.75 for single, double and triple charged ions respectively in the UniProt-database (version 56.0, human taxonomy (20069 entries)). Carboxymethylation of Cysteine (+57.021 u) as fixed and oxidation of Methionine (+15.996 u) as variable modifications and tryptic cleavage were considered. The number of allowed missed cleavages was 2, the mass tolerance for precursor ions was 10 ppm and for fragment ions 0.5 Da. The cut-off for mass differences with the theoretical mass of the identified peptides was set at 2 ppm.

Contamination of CSF by serum of plasma is a possible issue in CSF peptide profiling, because if one or more of the samples is contaminated by serum or plasma, the comparison of CSF peptide profiles is inevitably skewed by the higher total protein concentrations in serum or plasma [12]. To prevent inclusion of contaminated CSF samples in this study, the CSF samples were checked for specific blood contamination. If a hemoglobin peptide could be identified by nanoLC-ESI-Orbitrap (C_{18} column) with sufficiently high confidence score or if the mass peak 1274.7255 (part of hemoglobin gamma) has a signal to noise of 4 or higher in MALDI-FT-ICR measurements, the sample was discarded from further analysis due to plasma/serum contamination. Another blood specific protein, apolipoprotein B100 was checked in the same way as possible blood contamination.

Immuno-assays for validation of differentially abundant proteins

For two proteins, for which we found differentially abundant peptides, we performed validation experiments. This was done by commercially available ELISA (for vitamin D-binding protein) and by Western blot (for protein jagged-1), using 20 RR MScl and 20 PP MScl samples. These samples included the samples measured by mass spectrometry as well as an equal number of new CSF samples, for external validation. For the first protein, vitamin D-binding protein, we performed a commercially available ELISA (Immundiagnostik, Germany) according to the manufacturers specifications. For the second protein that was differentially abundant between the both MScl types, jagged-1, we performed a two-step western blot using goat anti-jagged1 antibodies (primary antibody) and anti-goat antibodies (secondary antibody) (Sigma Aldrich, United States). Protein transfer was checked by Ponceau staining. Quantitative assessment of the gel bands after photoluminescence was performed using Image J (freely available at www.rsb.info.nih.gov/ij).

RESULTS

Clinical information

In total 34 CSF samples were used for mass spectrometry analysis, while twenty of these samples and twenty additional samples were used for validation experiments. All samples analyzed by mass spectrometry were tested as being negative for serum/plasma contamination by MALDI-FT-ICR measurements. However, in three samples we were able to identify hemoglobin peptides with sufficiently high XC scores for confident identification using nanoLC-ESI-Orbitrap measurements. These three samples, 2 PP MScl and 1 RR MScl, were subsequently excluded from further analysis. The analysis matrix, which was used to profile the differences in peptide profile between the three groups, consisted therefore of 31 CSF samples (Table 1).

Table 1. CSF sample information. The concentration and age are averages with standard deviation in brackets. None of the variables in these tables differed significantly between the groups (all t-tests showed p-values higher than 0.05)

	PP MScl	RR MScl	Controls
Number of samples	10	11	10
Protein concentration (g/L)	0.398 (0.118)	0.391 (0.135)	0.386 (0.110)
Albumin concentration (g/L)	0.254 (0.104)	0.228 (0.082)	0.205 (0.090)
Age	48.1 (9.0)	43.9 (14.1)	5101 (13.7)
Male/Female ratio	6 / 4	6 / 5	8 / 2

Peak detection and data analysis

After the MALDI-FT-ICR spectra were loaded into Peptrix™ software package, they were each tagged with a group number (1, 2 and 3 for PP MScl, RR MScl and controls respectively). Calibrating using five omnipresent albumin peaks was followed by generation of an analysis matrix with the intensity of all peaks of every sample recorded for all detected peaks. Using the Wilcoxon-Mann-Whitney test to compare the groups pair-wise, the comparison between both MScl types resulted in 15 peak masses with p-values below 0.01. By scrambling the sample groups the number of background peaks was determined at 17, so the number of differentially abundant peptide peaks in the comparison of the two MScl types is around the level of the number of background peaks, indicating that the difference between these two groups appears to be nonexistent or at least at background level. However, the proteins that were identified with low p-values in this comparison were of substantial interest in a MScl context.

A total of 43 peptide peaks with a p-value below 0.01 were observed for the comparison of PP MScl versus the controls. The comparison of RR MScl versus the controls had 41 peak masses with p-values lower than 0.01. Seventeen of the peak masses with $p < 0.01$ were present in both comparisons.

Identification

Identification of the differentially abundant peptides was performed by measuring all samples using the nanoLC-ESI-Orbitrap. Due to the prefractionating by nanoLC far more peak masses and identifications are generated by ESI-Orbitrap than there are peak masses in the analysis matrix generated by quantitative MALDI-FT-ICR. Although many of the identified peptides do not correspond to peak masses in the analysis matrix, we were able to identify a number of differentially abundant peptides for all three comparisons (full list, including charge states and sequence coverage, can be found in the supplementary material). Of the 43 differentially abundant peptide masses that were observed using MALDI-FT-ICR mass spectrometry in the comparison of PP MScl versus the controls we were able to identify 28 peptides (Table 2). These peptides included several peptides of Ig gamma-1 and Ig kappa. Another differentially abundant peptide was identified as a part apolipoprotein D, which has previously been shown to be elevated intrathecally in MScl patients [13]. In the comparison of RR MScl versus controls we were able to identify 24 of the 41 differentially abundant peptide masses, which, as was the case with the comparison of PP MScl versus the controls, also included several peptides of Ig gamma-1 and Ig kappa as well as apolipoprotein D (Table 3). In fact, of the 24 peptides identified in this comparison, 14 were also identified in the comparison of PP MScl versus the controls, indicating that the differences of the two

Table 2. Differentially abundant peptides and proteins in the comparison of PP MScl versus controls.

p-value	Mass	Accession number	Protein	Peptide	Abundance in PP MScl	Fold increase	Incidence in	
							PP MScl (%)	Controls (%)
0.0008	1808.0093	P01857	Ig gamma-1 chain C region	VVSVLTVLHQDWLNGK	↑	3.036	100	90
0.0011	1677.8028	P01857	Ig gamma-1 chain C region	FNWYVDGVEVHNAK	↑	4.048	100	100
0.0017	1797.9237	Q2NKQ1	Small G protein signaling modulator 1	NTPTVLRPRDGSVDDR	↑	2.259	100	70
0.0032	1808.8544	O43707	Alpha-actinin-4	M*APYQGPDAVPGALDYK	↑	1.998	70	10
0.0032	2111.9749	Q8N8H1	KRAB domain-containing protein ZNF321	HESHHHIRDFCFQEIEK	↓	1.577	30	90
0.0032	2530.2363	P02787	Serotransferrin	SMGGKEDLIWELLNQAQEHFGK	↓	1.301	20	80
0.0034	2105.0748	Q8IZF0	Sodium leak channel non-selective protein	GKSLETLTQDHSNTVRYR	↑	1.625	80	20
0.0035	1797.8955	P01834	Ig kappa chain C region	SGTASVVCLNNFYPR	↑	2.761	100	100
0.0036	1740.8722	O43166	Signal-induced proliferation-associated 1-like protein 1	SQNGSLGSSVMAPVGPPR	↑	1.540	80	20
0.0038	1876.9188	Q96RD6	Pannexin-2	GGGGDPGPGPAPAPAPPPAPDK	↓	1.458	10	70
0.0038	2196.0455	A1L0T0	Acetolactate synthase-like protein	ENEDQVVKVLHDAQQQCR	↑	1.891	80	30
0.0042	1325.6870	Q9NW82	WD repeat-containing protein 70	KVIPTTCTYSR	↑	2.574	70	50
0.0044	1734.7859	Q96DR4	StAR-related lipid transfer protein 4	GYNHPCGWFCVPLK	↑	2.551	70	10
0.0044	2516.3324	Q1L5Z9	LON peptidase N-terminal domain and RING finger protein	TFPDGSSVVDAIGISRFRVLSHR	↑	2.847	50	20
0.0054	3270.4210	Q9UPA5	Protein bassoon	HSYHDYDEPPEEGLWPHDEG GPGRHASAK	↑	1.391	90	10
0.0054	2139.0232	P01857	Ig gamma-1 chain C region	TPEVTCVVVDVSHEDPEVK	↑	2.392	100	80
0.0054	1946.0258	P01834	Ig kappa chain C region	TVAAPSVFIFPPSDEQLK	↑	2.478	100	80
0.0054	2657.1775	Q9PKX4	Docking protein 6	M*CDTGEGLFTFQTRREGEMIYQK	↑	1.191	70	10
0.0054	1423.7421	P05090	Apolipoprotein D	NPNLPPETVDSLK	↑	2.579	100	50
0.0063	1681.8125	Q9BZ29	Dedicator of cytokinesis protein 9	YAYKAEPYVASEYK	↑	2.794	70	20
0.0063	1682.8158	Q9Y2P8	RNA 3′-terminal phosphate cyclase-like protein	GMAYSVRVSPQM*ANR	↑	1.922	70	10
0.0064	1683.8156	P21817	Ryanodine receptor 1	EIRFPKMVTSCCR	↑	2.129	70	10
0.0069	1639.7817	P02768	Albumin	DVFLGM*FLYEYAR	↑	1.732	100	80
0.0069	1865.0194	P01614	Ig kappa chain V-II region Cum	LEIPYTFGQGTKLEIR	↑	1.376	70	10

Table 2 (continued)

p-value	Mass	Accession number	Protein	Peptide	Abundance in PP MScl	Fold increase	Incidence in	
							PP MScl (%)	Controls (%)
0.0069	2747.3214	Q7L3B6	Hsp90 co-chaperone Cdc37-like 1	MCLWSTDAISKDVFNKSFINQDK	↑	1.719	80	20
0.0071	1147.5513	Q9NRL2	Bromodomain adjacent to zinc finger domain protein 1A	QEQINCVTR	↓	1.039	30	100
0.0082	1163.5746	Q5TG30	Rho GTPase-activating protein 18-like	IGDLSLQDM*R	↓	1.124	20	80
0.0092	1908.9722	O75746	Calcium-binding mitochondrial carrier protein Aralar1	EEGPSAFWKGTAARVFR	↑	1.513	40	50
0.0092	1268.6771	Q8NEB9	Phosphatidylinositol 3-kinase catalytic subunit type 3	SALM*PAQLFFK	↓	-	0	60

Table 3. Differentially abundant peptides and proteins in the comparison of RR MScl versus controls.

p-value	Mass	Accession number	Protein	Peptide	Abundance in RR MScl	Fold increase	Incidence in	
							RR MScl (%)	Controls (%)
0.0016	1797.8955	P01834	Ig kappa chain C region	SGTASVVCLNNFYPR	↑	2.790	100	80
0.0016	1797.9237	Q2NKQ1	Small G protein signaling modulator 1	NTPTVLRPRDGSVDDR	↑	2.417	100	70
0.0032	1734.7859	Q96DR4	StAR-related lipid transfer protein 4	GYNHPCGWFCVPLK	↑	2.837	27	10
0.0032	1865.0194	P01614	Ig kappa chain V-II region Cum	LEIPYTFGQGTKLEIR	↑	1.678	73	10
0.0038	1808.0093	P01857	Ig gamma-1 chain C region	VVSVLTVLHQDWLNGK	↑	2.418	100	70
0.0041	1325.6870	Q9NW82	WD repeat-containing protein 70	KVIPTTCTYSR	↑	3.054	82	50
0.0043	1740.8722	O43166	Signal-induced proliferation-associated 1-like protein 1	SQNGSLGSSVMAPVGPPR	↑	1.482	64	20
0.0047	1946.0258	P01834	Ig kappa chain C region	TVAAPSVFIFPPSDEQLK	↑	2.967	100	100
0.0047	1186.6443	P01857	Ig gamma-1 chain C region	GPSVFPLAPSSK	↑	3.872	100	100
0.0047	1126.5756	Q13555	Calcium/calmodulin-dependent protein kinase type II gamma chain	KTSTQEYAAK	↓	1.413	27	80
0.0055	2803.2526	Q12772	Sterol regulatory element-binding protein 2	LPAGSACSDVHMALCAVN LAECAEEK	↑	1.284	64	80
0.0051	1423.7421	P05090	Apolipoprotein D	NPNLPPETVDSLK	↑	2.824	82	50

Table 3 (continued)

p-value	Mass	Accession number	Protein	Peptide	Abundance in RR MScl	Fold increase	Incidence in	
							RR MScl (%)	Controls (%)
0.0058	2516.3324	Q1L5Z9	LON peptidase N-terminal domain and RING finger protein	TFPDGSSVVDAIGISRFRVLSHR	↑	2.212	73	20
0.0058	1859.9226	P07093	Glia-derived nexin	VLGITDM*FDSSKANFAK	↑	2.097	73	10
0.0058	2139.0232	P01857	Ig gamma-1 chain C region	TPEVTCVVVDVSHEDPEVK	↑	2.698	100	80
0.0060	1682.8158	Q9Y2P8	RNA 3′-terminal phosphate cyclase-like protein	GMAYSVRVSPQM*ANR	↑	2.112	64	10
0.0062	1996.8241	Q9NXT0	Zinc finger protein568	DQGGHSGERPYECGEYR	↑	1.969	82	30
0.0063	1683.8156	P21817	Ryanodine receptor 1	EIRFPKMVTSCCR	↑	1.877	55	10
0.0065	1605.8326	Q86VR8	Four-jointed box protein 1	SEPRWHVSARQPR	↓	-	0	60
0.0065	2576.3346	Q19AV6	Zinc finger SWIM domain-containing protein 7	HLLAVYLSQVMRTCQQSVSDk	↓	1.318	36	20
0.0065	2736.3327	P12931	Proto-oncogene tyrosine-protein kinase Src	TQFNSLQQLVAYYSKHADGLCHR	↑	1.329	73	10
0.0070	2045.1381	O15033	Protein KIAA0317	EFYLKIIPWRLYTFR	↑	1.445	100	80
0.0071	1677.8028	P01857	Ig gamma-1 chain C region	FNWYVDGVEVHNAK	↑	3.804	91	90
0.0086	1275.6857	P01009	Alpha-1-antitrypsin	GKWERPFEVK	↑	1.444	100	100

MScl types compared to the controls are remarkably similar. Additional proteins with low p-values in the comparisons of both MScl types with the controls include StAR-related lipid transfer protein 4 (2.551 fold increase in PP MScl compared to controls and 2.837 fold increase in RR MScl compared to controls), LON peptidase N-terminal domain and RING finger protein (2.847 fold increase in PP MScl and 2.212 fold increase in RR MScl), and ryanodine receptor 1 (2.129 fold increase in PP MScl and 1.877 fold increase in RR MScl).

PP MScl versus RR MScl

The comparison of PP MScl versus RR MScl showed a limited number of differentially abundant peptide peaks. Of these peaks 7 were identified, the most notable being protein jagged-1 (Table 4). This particular protein was over three times less abundant in PP MScl compared to RR MScl. Another interesting differentially abundant protein is vitamin D-binding protein, which was not present in the PP MScl samples but was present in the RR MScl samples. The number of peaks with low p-values in the comparison of the two MScl types is lower than the number

Table 4. Differentially abundant peptides and proteins in the comparison of PP MScl versus RR MScl.

p-value	Mass	Accession number	Protein	Peptide	Abundance in PP MScl	Fold increase	Incidence in	
							PP MScl (%)	RR MScl (%)
0.0014	1148.5563	P61769	Beta-2-microglobulin	VEHSDLSFSK	↑	-	70	0
0.0015	1268.6771	Q8NEB9	Phosphatidylinositol 3-kinase catalytic subunit type 3	SALM*PAQLFFK	↓	-	0	73
0.0041	2071.7953	Q9UBE8	Serine/threonine kinase NLK	YHTCM*CKCCFSTSTGR	↑	-	60	0
0.0069	2105.0748	Q8IZF0	Sodium leak channel non-selective protein	GKSLETLTQDHSNTVRYR	↑	1.180	80	18
0.0071	1996.8241	Q9NXT0	Zinc finger protein568	DQGGHSGERPYECGEYR	↓	1.786	80	82
0.0087	2143.8653	P78504	Protein jagged-1	TCMEGWM*GPECNRAICR	↓	3.188	30	63
0.0092	2213.1255	P02774	Vitamin D-binding protein	ELPEHTVKLCDNLSTKNSK	↓	-	0	55

of background peaks, so this is a strong indication that, even though there may be peptides and proteins that are differentially abundant in the comparison, overall the difference between the two disease types appears to be undetectable by statistical means. Other proteins, such as serine/threonine kinase NLK and sodium leak channel non-selective protein were more abundant in PP MScl than in RR MScl, although for the latter protein the difference was small (1.18 fold increase).

Because our main interest was focused on the differences between the two MScl types we selected two differentially abundant proteins from that comparison for validation purposes using the 10 PP MScl and 10 RR MScl samples measured by mass spectrometry and an additional 10 PP MScl and 10 RR MScl samples. Validation by ELISA showed that the concentration of vitamin D-binding protein was significantly lower (t-test, $p = 0.0058$) in the PP MScl group compared to the RR MScl group (table 5). After Western blotting, quantitative assessment of the gel bands showed that protein jagged-1 was indeed less abundant in PP MScl than in RR MScl. A t-test on the photoluminescence readout values of the original sample set showed a p-value of 0.0173 when comparing the RR and PP MScl samples; the same comparison in the combined data set of the original samples with the validation samples had a p-value of 0.0162, indicating a significant differential abundance of jagged-1 in the two MScl disease types (table 5).

Table 5. The results of the validation experiments. By ELISA measurement vitamin D binding protein is more abundant in RR MScl than in PP MScl in the sample set in which the original samples were combined with an equal number of validation samples (p = 0.0058), based on the average (+/- standard deviation) concentrations in CSF. Protein jagged-1 (Western blot) is more abundant in RR MScl than in PP MScl in the same sample set mentioned above (p = 0.0162), based on the averages (+/- standard deviation) in photoluminescence readout.

	PP MScl		RR MScl		
	Average	Standard deviation	Average	Standard deviation	p-value (t-test)
Vitamin D-binding protein concentration in pg/ml (ELISA)	14950	5985	20973	7013	0.0058
Protein jagged-1 photoluminescence readout (Western blot)	9678	5153	17332	11900	0.0124

DISCUSSION

The main finding of this comparative study is the observation that the proteome profiles of CSF in PP vs RR MScl patients overlap to a large extent. This is in line with the lack of clear-cut differences between the two major clinical MScl sub-groups, at genetic, immunological and neuropathological levels. Interestingly, our approach using sensitive state-of the art mass spectrometry techniques, led to the identification of a few distinct CSF proteins, some of them with biological functions that appear of direct interest for MScl pathology.

Because of the healthy state of the CSF control group, no intrathecal inflammatory response was to be expected in this group. Therefore the very clear difference in immunoglobulin abundance in the comparison with the both MScl disease types can be explained. Previous studies included a large number of different diseases in the control groups, including inflammatory CNS diseases, which influenced results. Whereas chromogranin A, complement C3 and clusterin were significantly differentially expressed in the previous two MScl proteomics studies that we have published [8, 10], this was not the case here. Although the previously differentially identified peptides of these three proteins did have p-values approaching statistical significance (p = 0.046, p = 0.059 and p = 0.038 for chromogranin A, complement C3 and clusterin, respectively), none of them was below the cut-off value of p = 0.01. However, if the two groups of MScl subtypes are combined and then compared to the controls, the p-values for the previously identified peptides of chromogranin A, complement C3 and clusterin are 0.009, 0.013 and 0.008, respectively.

Although the number of differentially abundant peptide peaks in the comparison of both MScl types is in the same range as the statistical background, some differences can still be clearly observed. One of the peptides of protein jagged-1 is over three times less abundant in PP MScl and also is observed with a lower incidence in this group (Table 4). Protein jagged-1 is a ligand for multiple Notch recep-

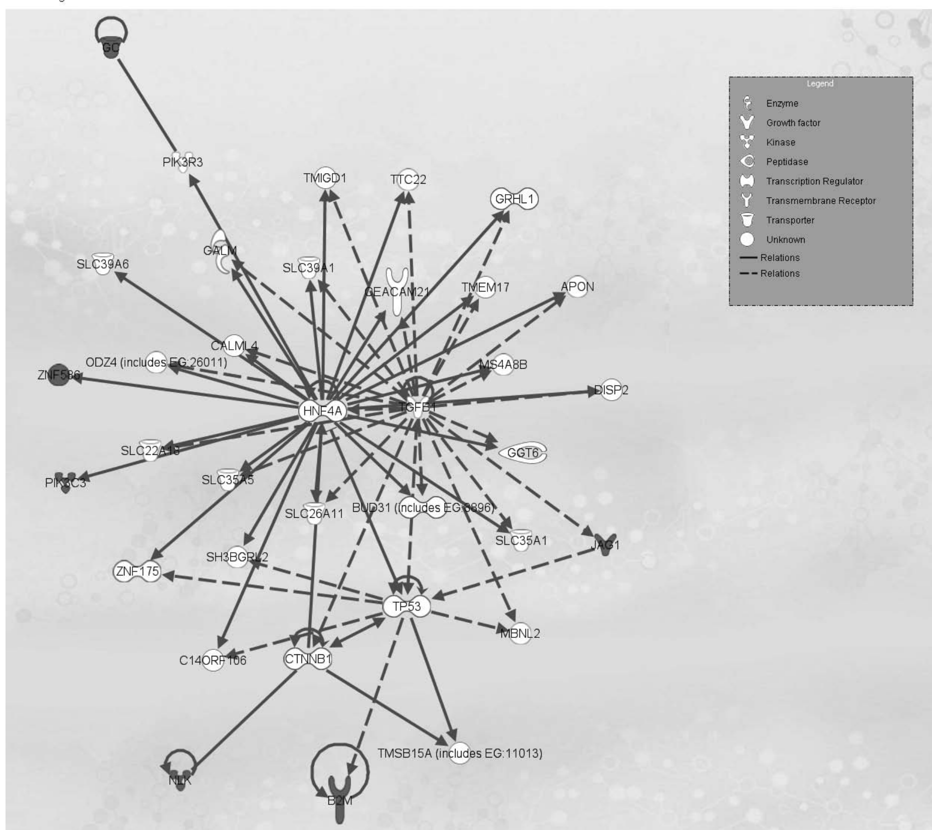
tors and involved in the mediation of Notch signaling, which influences neuronal function and development [14]. The Notch signaling pathway has long been known to influence cell fate in the developing nervous system. Jagged-1 has been found to be highly expressed in hypertrophic astrocytes within and around active MScl plaques lacking remyelination, while, in contrast, there was negligible jagged-1 expression in remyelinated lesions suggesting involvement of the Notch pathway in remyelination in MScl [15]. Later, linkage equilibrium screening implicated a number of genes, including the jagged-1 gene, as susceptibility genes for MScl in a large contingent of Europeans [16]. It has also been suggested that jagged-1 has therapeutic potential in the treatment of $CD8^+$ T cell mediated diseases, due to it's ability to deliver indirect negative signals into $CD8^+$ T cells *in vivo* [17]. Additionally, animal models have shown that elevated expression of Notch and jagged-1 expression does not appear to be a limiting factor in remyelination, but the animal model study reports that there were no quantitative differences in Notch1 expressing cells in slow and rapidly remyelinating lesions, indicating that Notch–Jagged signaling is not a rate-limiting determinant of remyelination in rodent models of demyelination [18]. Additionally immunohistochemistry experiments have shown that constituents of the Notch pathway are expressed in remyelination in an animal model of T-cell- and antibody-mediated CNS demyelination [19]. However, network studies based on the quantitative expression levels of 20 genes in over one hundred individuals identify jagged-1 as a new therapeutic target whose differential behavior in the MScl network was not modified by immunomodulatory therapy, illustrating how network analysis can predict therapeutic targets for immune intervention and identifying the immunomodulatory properties of jagged-1, making it a new therapeutic target for MScl and other autoimmune diseases [20]. The identified peptide of vitamin D-binding protein is not observed in any of the PP MScl samples, but small peaks of this peptide are detected in 6/11 RR MScl samples. Impaired vitamin D homeostasis has been widely implicated in MScl for some years now [21-23]. This vitamin directly and indirectly regulates the differentiation, activation of CD4+ T-lymphocytes and can prevent the development of autoimmune processes [24, 25], and so it may be involved MScl. Considering that the geographic incidence of MScl indicates an increase in MScl with a decrease in sunlight exposure, that vitamin D is produced in the skin by solar or UV irradiation and that high serum levels of 25-hydroxyvitamin D have been reported to correlate with a reduced risk of MScl, a protective role of vitamin D has been suggested [26]. More than 99% of 25-hydroxyvitamin D, the principle circulating metabolite of vitamin D is bound to proteins, of which approximately 90% is bound to vitamin D-binding protein [27, 28]. Recently, a CSF proteomics study showed that serum levels of vitamin D-binding protein were decreased significantly in RR MScl patients

compared to other neurological disorders [29]. While our results do not indicate a differential abundance difference between the MScl subtypes and the controls, the two disease types did vary significantly in CSF levels of vitamin D-binding protein, with RR MScl showing a higher abundance. Since a neuroprotective function has been suggested for vitamin D, it may be that in PP MScl this neuroprotective pathway is at least partially deficient, resulting in a significantly more disabling disease manifestation. However, it should not be forgotten that vitamin D-binding protein has pleiotropic functions, beyond vitamin D metabolism. It can significantly enhance the chemotactic response to complement fragment C5a [30], and there are substantial stimulatory effects on macrophages [31]. In light of the increasingly recognized role of innate immunity in the progressive phase of MScl pathogenesis [32], vitamin D-binding protein appears an interesting candidate mediator.

The identification of a peptide of beta-2-microglobulin as differentially abundant in the comparison between both MScl disease types is somewhat misleading. Another seven peptides of this protein were identified among the peak masses in the analysis matrix that had high p-values in the comparison of PP and RR MScl, so it is very likely that the low p-value of the peptide of beta-2-microglobulin reflects high abundant protein variations, suggesting this low p-value is most likely a false positive. In contrast, the other proteins that are differentially abundant in this comparison were either identified by the single peptide listed in table 4 or by multiple peptides that had low p-values.

In order to place the identified proteins in a biological context, they were uploaded to the Ingenuity Pathways Analysis service (Ingenuity Systems) for to network analysis. Six of the seven differentially abundant proteins were placed in a network relating to neurological disease (figure 1).

In the comparisons of both MScl types with the controls, several proteins are present in both comparisons, for example Ig gamma-1 chain C region and apolipoprotein D. Apolipoproteins have been previously implicated in MScl. Proteomics studies have shown apolipoprotein E abundances to be elevated in CSF of MScl patients compared to controls [7, 10]. Apolipoprotein D is a lipocalin that is, in the central nervous system, mainly expressed in glia, but also in neurons. This protein has been repeatedly implicated in MScl, and it has been shown that it has a neuroprotective effect in a number of neurodegenerative diseases by controlling the level of peroxidated lipids, which coincides with glial activation in mouse models of encephalitis [33]. A previous proteomics study showed increased levels of apolipoprotein D in patients with a clinically isolated syndrome of demyelination, indicating that abundance levels of this protein are highest in MScl patients at the time of their first exacerbation [8, 13]. Another potentially interesting protein found to be differentially elevated in the comparisons of both MScl types with the



© 2000-2009 Ingenuity Systems, Inc. All rights reserved.

Figure 1. Six of the seven differentially abundant proteins identified in the comparison of the two MScl disease types (PP and RR) fit into a network related to neurological disease.

controls is ryanodine receptor 1. This receptor is involved in the maintenance of the calcium-equilibrium in brain tissue. The release of toxic levels of positively charged calcium ions may, due to the deleterious effects of excitotoxicity, represent a key mechanism of axonal degeneration in disorders such as MScl [34].

The identification of peptides of albumin and serotransferrin stands out in the tables of the differentially abundant peptides and proteins. While the identifications are essentially correct it must be noted that only 2.4% of the identified albumin peptides and 4.8% of the identified serotransferrin peptides had p-values below 0.01, which means that the values for the peptides of these two particular proteins are most likely due to other reasons than large abundance differences in these two proteins. In comparison, most of the other differentially abundant proteins that we identified were found by a small number of peptides with a low p-value or a single peptide with a low p-value. For these proteins we did not observe any other pep-

tides with non-significant p-values, with the exception of bromodomain adjacent to zinc finger domain protein 1A and Rho GTPase-activating protein 18-like. For both of these proteins another peptide was identified with high p-value, indicating these proteins were likely not differentially abundant. The peptides with low p-values of these proteins were also characterized by a low fold increase, making them less interesting for independent immunoassay follow-up.

In conclusion, the CSF peptide profile of the control samples clearly differed from both MScl types, with, not unexpectedly, proteins related to immune response showing the highest fold increase in abundance in the MScl types compared to the controls. Even though the CSF peptide profiles measured by MALDI-FT-ICR of PP MScl and RR MScl were quite similar, still a few differences could be observed, most notably regarding the molecules confirmed by immunoassay, protein jagged-1 and vitamin D-binding protein.

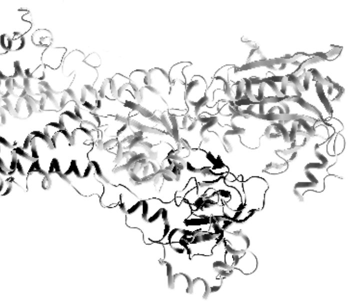
ACKNOWLEDGEMENTS

This work was supported by a program grant provided by the Netherlands Multiple Sclerosis Research Foundation to the Rotterdam Multiple Sclerosis research Center, ErasMS, the Netherlands Organization for Scientific research (ZON-MW, RQH), Hersenstichting Nederland, the Top Institute Pharma grant D4-102-1 and the UEPHA-MS grant PITN-GA-2008-212877.

REFERENCES

1. Compston, A., et al., *Multiple sclerosis.* Lancet, 2002. **359**: p. 1221-31.
2. Miller, D.H., et al., *Primary-progressive multiple sclerosis.* Lancet Neurol, 2007. **6**: p. 903-12.
3. Hoppenbrouwers, I.A., et al., *Familial clustering of multiple sclerosis in a Dutch genetic isolate.* Mult Scler, 2007. **13**: p. 17-24.
4. Richards, R.G., et al., *A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.* Health Technol Assess, 2002. **6**: p. 1-73.
5. Lassmann, H., *New concepts on progressive multiple sclerosis.* Curr Neurol Neurosci Rep, 2007. **7**: p. 239-44.
6. Han, M.H., et al., *Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets.* Nature, 2008. **451**: p. 1076-81.
7. Lehmensiek, V., et al., *Cerebrospinal fluid proteome profile in multiple sclerosis.* Mult Scler, 2007. **13**: p. 840-9.
8. Stoop, M.P., et al., *Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry.* Proteomics, 2008. **8**: p. 1576-85.
9. Chiasserini, D., et al., *CSF proteome analysis in multiple sclerosis patients by two-dimensional electrophoresis.* Eur J Neurol, 2008. **15**: p. 998-1001.
10. Stoop, M.P., et al., *Quantitative matrix-assisted laser desorption ionization-fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple-sclerosis-related proteins.* J Proteome Res, 2009. **8**: p. 1404-14.
11. Olsen, J.V., et al., *Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.* Mol Cell Proteomics, 2005. **4**: p. 2010-21.
12. You, J.S., et al., *The impact of blood contamination on the proteome of cerebrospinal fluid.* Proteomics, 2005. **5**: p. 290-6.
13. Reindl, M., et al., *Increased intrathecal production of apolipoprotein D in multiple sclerosis.* J Neuroimmunol, 2001. **119**: p. 327-32.
14. Gaiano, N., et al., *The role of notch in promoting glial and neural stem cell fates.* Annu Rev Neurosci, 2002. **25**: p. 471-90.
15. John, G.R., et al., *Multiple sclerosis: re-expression of a developmental pathway that restricts oligodendrocyte maturation.* Nat Med, 2002. **8**: p. 1115-21.
16. The Games Collaborative, G., et al., *Linkage disequilibrium screening for multiple sclerosis implicates JAG1 and POU2AF1 as susceptibility genes in Europeans.* J Neuroimmunol, 2006. **179**: p. 108-16.
17. Kijima, M., et al., *Jagged1 suppresses collagen-induced arthritis by indirectly providing a negative signal in CD8+ T cells.* J Immunol, 2009. **182**: p. 3566-72.
18. Stidworthy, M.F., et al., *Notch1 and Jagged1 are expressed after CNS demyelination, but are not a major rate-determining factor during remyelination.* Brain, 2004. **127**: p. 1928-41.
19. Seifert, T., et al., *Notch1 and its ligand Jagged1 are present in remyelination in a T-cell- and antibody-mediated model of inflammatory demyelination.* Acta Neuropathol, 2007. **113**: p. 195-203.
20. Palacios, R., et al., *A network analysis of the human T-cell activation gene network identifies JAGGED1 as a therapeutic target for autoimmune diseases.* PLoS ONE, 2007. **2**: p. e1222.
21. Cantorna, M.T., *Vitamin D and autoimmunity: is vitamin D status an environmental factor affecting autoimmune disease prevalence?* Proc Soc Exp Biol Med, 2000. **223**: p. 230-3.

22. Lips, P., *Which circulating level of 25-hydroxyvitamin D is appropriate?* J Steroid Biochem Mol Biol, 2004. **89-90**: p. 611-4.
23. Szodoray, P., et al., *The complex role of vitamin D in autoimmune diseases.* Scand J Immunol, 2008. **68**: p. 261-9.
24. Adorini, L., *Selective immunointervention in autoimmune diseases: lessons from multiple sclerosis.* J Chemother, 2001. **13**: p. 219-34.
25. Cantorna, M.T., et al., *Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence.* Exp Biol Med (Maywood), 2004. **229**: p. 1136-42.
26. Raghuwanshi, A., et al., *Vitamin D and multiple sclerosis.* J Cell Biochem, 2008. **105**: p. 338-43.
27. Bikle, D.D., et al., *Serum protein binding of 1,25-dihydroxyvitamin D: a reevaluation by direct measurement of free metabolite levels.* J Clin Endocrinol Metab, 1985. **61**: p. 969-75.
28. Bikle, D.D., et al., *Assessment of the free fraction of 25-hydroxyvitamin D in serum and its regulation by albumin and the vitamin D-binding protein.* J Clin Endocrinol Metab, 1986. **63**: p. 954-9.
29. Qin, Z., et al., *Alteration of DBP levels in CSF of patients with MS by proteomics analysis.* Cell Mol Neurobiol, 2009. **29**: p. 203-10.
30. Shah, A.B., et al., *Selective inhibition of the C5a chemotactic cofactor function of the vitamin D binding protein by 1,25(OH)2 vitamin D3.* Mol Immunol, 2006. **43**: p. 1109-15.
31. Gomme, P.T., et al., *Therapeutic potential of vitamin D-binding protein.* Trends Biotechnol, 2004. **22**: p. 340-5.
32. Weiner, H.L., *The challenge of multiple sclerosis: how do we cure a chronic heterogeneous disease?* Ann Neurol, 2009. **65**: p. 239-48.
33. Do Carmo, S., et al., *Neuroprotective effect of apolipoprotein D against human coronavirus OC43-induced encephalitis in mice.* J Neurosci, 2008. **28**: p. 10330-8.
34. Ouardouz, M., et al., *Glutamate receptors on myelinated spinal cord axons: I. GluR6 kainate receptors.* Ann Neurol, 2009. **65**: p. 151-9.



Chapter 7
Discussion

In this thesis we describe the mass spectrometry experiments that were performed on cerebrospinal fluid (CSF) during our efforts to identify biomarkers for elucidation of disease pathology for multiple sclerosis (MScl). The actual experimental work is preceded by a review of the current knowledge of CSF biomarkers for the disease. In the experiments described in this thesis an effort was made to catalogue the inter-individual variation of protein and metabolite levels in CSF of neurologically normal control patients. Also, the applied methodology was critically examined by performing experiments to pinpoint the determining factors in peptide and protein database searching. And finally, three different biomarker discovery studies were performed on CSF of patients with MScl, including an effort to differentiate between two different clinical subtypes of MScl. The pathology of MScl, a highly heterogeneous disease of the central nervous system (CNS), is characterized by a combination of factors such as inflammation, demyelination and axonal damage [1, 2]. A number of different techniques and research fields are employed in current efforts to elucidate MScl disease pathology.

Magnetic Resonance Imaging (MRI) is playing an increasing role in the scientific investigation and clinical management of MScl [3]. Dissemination of MScl plaques in time and space, visualized by MRI, are included in the criteria for clinical diagnosis of MScl [4]. Although conventional MRI assessment of lesions provide an important tool to monitor disease course, its limitations include a weak association with clinical status and a lack of sensitivity to other clinically relevant findings [5-7]. New developments, such as contrast agents composed of iron particles that can be used to track macrophages or non-conventional MRI techniques, can be employed for a more in-depth MRI analysis [8-10]. However, even though these techniques provide intriguing results about brain damage in MScl, the main challenge here remains the establishment of a better correlation between clinical and MRI findings.

Whereas MRI based studies of MScl are used to monitor effects of the disease on the brain and the spinal cord, other approaches are more focused on the specific disease components like demyelination or inflammation. Immunology based studies on lymphocytes, for example auto-reactive $CD8^{+}$ T cells, could provide important clues about the pathology of autoimmunity in MScl and perhaps even provide new therapeutic targets [11]. Due to their capacity to produce toxic inflammatory mediators, macrophages are believed to play a pivotal role in the CNS in MScl [12, 13]. However, not all macrophages show pro-inflammatory effects, as foamy macrophages consistently express a series of anti-inflammatory molecules [14]. So not all components of the immune response in MScl are by definition negative effects. And understanding the mechanisms behind naturally occurring counter-regulatory processes may also allow for new cellular therapy targets.

Another interesting topic in MScl is genetics. Genetic complexity and interactions with unidentified environmental factors are current barriers for the elucidation of the etiology of the disease [15]. But the fact that there appears to be a modest but significant concordance for disease course in siblings [16], and that genetic variation in specific loci in isolated populations influences MScl susceptibility [17], does certainly strongly suggest a genetic factor in the disease. Results from whole genome association studies support, at least partly, an immunological basis for the disease [18].

CURRENT MULTIPLE SCLEROSIS BIOMARKERS

Despite extensive literature on CSF biomarkers in MScl only qualitative and quantitative methods for determining the intrathecal production of immunoglobulins are used as an aid to clinical diagnosis of MScl [19]. Since MScl lesions are rarely biopsied, CSF is the closest one can get to the pathology of MScl and is therefore often referred to as being a "liquid biopsy". Unfortunately, the majority of MScl CSF biomarker studies are not performed and reported in a standardized way and published results lack sufficient detail to allow a critical review of the study and reproduction of the study design [20]. In diseases with a complex pathogenesis, such as MScl, an individual biomarker is likely to reflect only one of the many ongoing pathogenic processes. Since MScl is comprised of both a neuroinflammatory as well as a neurodegenerative component it seems logical to divide the biomarkers for MScl using a similar division. MScl related inflammatory biomarkers correlate with MRI markers of disease activity [21]. When compared to control patients with non-inflammatory diseases of the CNS, subjects with MScl show clear evidence of intermittent inflammatory activity in both the CNS and in the periphery [22, 23], and differences appear to occur between different MScl disease types [24, 25]. Inflammatory markers include, for example, an increase in pro-inflammatory cytokines and an increase in interleukin-2 receptors in CD4+ T-helper cells [26, 27]. Markers for neurodegeneration include markers for demyelination and axonal loss. During demyelination, myelin basic protein and its fragments are released into the CSF and can be used as an index of active demyelination [28], but this has never been developed for clinical practice. CSF concentrations of actin, tubulin and neurofilament light chain, structural neuron-specific protein, correlate with disability [29], which is reflected by the fact that CSF neurofilament light chain concentrations are particularly high soon after onset of an attack and decrease with time [30]. Tau, or microtubule-associated phosphoprotein is found predominantly within axons, and it has been reported that tau levels were increased in MScl

compared to controls and the levels of this protein were highest in patients with progressive disease [31]. However, other studies noted large variations and no increased CSF tau levels in MScl compared to controls [32, 33]. Other molecules, such as the neuron-specific molecule N-acetylaspartate, are reported to decrease as the disease in concentration progresses [34].

PROTEOMICS-BASED BIOMARKER DISCOVERY

The close proximity of CSF to the disease processes of MScl makes it an interesting body fluid in which to search for biomarkers and disease-associated proteins and peptides [35, 36]. Due to the complexity and heterogeneity of MScl it is unlikely that a single biomarker will be found that either reflects all ongoing pathological processes or completely and accurately predicts the disease progression in individuals. An advanced, unbiased discovery technique like proteomics is able to measure large numbers of samples in relatively short periods of time. These types of experiments yield large amounts of data which, when statistically analyzed, could play a significant role in the detection of novel candidate biomarkers. CSF proteomics is a rapidly developing area of MScl biomarker research that, although still relatively new, has shown some interesting discoveries and will develop further to handle the huge protein complexity in clinical samples.

Essentially, every proteomics based biomarker discovery study should be comprised of the same five steps; standardized sample collection, sample pre-treatment, peptide profiling, peptide identification and validation. Although the first two steps seem fairly simple, they are just as essential to a good study as the latter three steps. As CSF sampling is not without risk, standardization of the methods used is beneficial for both the reproducibility of the procedure, and subsequently its analysis results, as well as the safety of the patients. Consequently, successful CSF sampling should meet a number of well defined criteria [37]. Any CSF sample containing blood cells or any other indication of blood contamination should be excluded from analysis because it does not accurately portray the real biological situation in the CSF of patients, due to the 400-fold higher total protein concentration in blood compared to CSF [38, 39].

One of the main sources of technical or sample variation is the pre-analytical phase [40]. Lack of standardized procedures for sample acquisition, handling and storage can account for a great part of all errors within the entire diagnostic process [41]. For successful generation of validated biomarkers, one must be aware of these potential pre-analytical stage variables. This is even more important if these studies are collaborations of multiple centres, since there is no standardized

global method of bio-fluid sampling for biomarker research [42]. In the studies presented in this thesis, all samples originated from the same university hospital and sample acquisition, using a single sampling method, was supervised by an experienced neurologist, allowing us to conclude that the pre-analytical variation in sample collection was as limited as possible. Also, since long-term sample storage at higher temperatures can influence sample composition, all samples were stored at –80°C [43].

At present, there are two dominant approaches for large-scale proteome analysis. The first is based on protein separation by 2D-gels, which is hampered by limited throughput, whilst the second, shotgun proteomics, excels in high-throughput applications [44]. However, proteome samples become much more complex due to the digestion of proteins, as one protein can generate dozens of peptides. As a result, the complexity of proteome samples is far beyond the capacity of modern analytical systems and the large dynamic range in protein abundance exceeds the dynamic range of all available analytical platforms. Several forms of sample pre-treatment can be used to reduce the complexity of samples. Two of the more commonly used methods are depletion of high abundant proteins and targeted enrichment of specific proteins of interest [45, 46]. The key to these approaches is to significantly reduce the sample complexity and meanwhile keeping the integrity of the proteome intact, since the fact remains that an increase of sample pre-treatment steps also increases the change of variation between the samples due to the pre-treatment. In order to limit these variations, the sample pre-treatment steps in the studies presented in this thesis were limited to a simple enzymatic digestion by trypsin, followed by either a desalting step for Matrix-Assisted Laser Desorption/Ionization (MALDI) mass spectrometry techniques or high resolution nanoLC separation for Electrospray Ionization (ESI) mass spectrometry. Although this approach limits the analysis, since the very low abundant proteins are not observed using this approach, it does allow for confident analysis of the high-to-low abundant proteins, since the lack of extended pre-treatment steps ascertains a very limited extra variation due to sample handling.

BIOLOGICAL VARIATION

One of the main challenges of biomarker discovery in body fluids is the large biological variation between individuals. Although measurement of large numbers of samples can be used to overcome biological, and also technical variation, this does require fast and reliable screening techniques [47]. Although shotgun proteomics is such a technique, it is important to investigate what type of variations can be

expected in a specific body fluid. In Chapter 2 the variations between individuals in abundance of 130 proteins in CSF are discussed.

Here we examined nine well-defined non-neurological CSF samples and found profound differences in protein abundances between individuals. Characterization of variation of CSF levels of amyloid beta [48] and apolipoprotein E [49] in patients with Alzheimer's disease have been published, but this is the first attempt to characterize a large number of proteins in CSF of patients without neurological afflictions. Some proteins, such as for example vitamin D-binding protein and transthyretin, appear to be more constant than others with regards to abundance levels in CSF. Other proteins, such as contactin-2 and haptoglobin, showed large variations in abundance levels. Since the variation of protein levels between individuals appears to be extensive as well as dependent on the proteins itself, it seems apparent that knowledge of the range of inter-individual variation of protein levels in CSF is an important instrument for assessment of potentially elevated protein levels when comparing a group of CSF samples from patients to a group of CSF controls. As a result, proteins showing high variation between individuals ought to be assessed with caution as candidate biomarkers because at least part of the difference observed between the diseased individuals and the controls will not be caused by the disease, but rather by the natural biological variation between individuals.

TECHNICAL VARIATION

Another issue that must be accounted for is the technical variance of the applied measurement method. This, like the biological variation, can be overcome by measuring a large enough number of samples [50-52]. The size and numbers of the differences observed between the analyzed groups must be confidently not due to technical variance, and this knowledge is based largely on the number of samples analyzed [53]. Another way to compensate for technical variation is provided by novel technological advances made in the mass spectrometry research field. For example, the reproducibility of peak heights in MALDI-Time-of-Flight (TOF) mass spectrometry is relatively poor (CV = 30-40%) [54], which makes quantitative MALDI-TOF experiments technically almost impossible. But newer MALDI techniques, such as MALDI-Fourier Transform Ion Cyclotron Resonance (FT-ICR) or MALDI triple quadrupole mass spectrometry shows far lower CV's of approximately 10%, which certainly allows for quantitative peptide profiling [55, 56].

Proteome analysis of clinical samples is a multidimensional assay that encompasses the comparative analysis of large numbers of variables that exceed the

number of samples to be analyzed by orders of magnitude [57]. Also, to obtain statistically significant data, the increasing number of analyzed components requires increasing the number of analyzed samples and, consequently, the need for greater computing power. Therefore, it is important to find a balance between the desire for maximal data for analysis and the amount of effort and analysis time needed [58]. In our studies we elected to aim for maximal data output, limiting the studies only to the number of samples we could collect for each group, in order to obtain the maximum amount of information from these experiments.

After acquisition of the mass spectra and statistical evaluation, shotgun proteomics yields a list of differentially abundant peptide masses relating to peptides and proteins that are differentially abundant in the sample groups. To identify these peptide masses MS/MS spectra are acquired of these peptide masses and these are subsequently database-searched using a software package such as Mascot [59] or SEQUEST [60]. Although these database searching software packages are highly evolved, the approach remains probability based, thus potential misidentifications cannot be excluded. If different mass spectrometers are used, extra attention ought to be given to false discovery. As discussed in Chapter 3, we used fifteen CSF samples to determine the number of false positive matches in sequence of peptide identification using different mass spectrometers (MALDI-TOF/TOF, MALDI-FT-ICR and ESI-Orbitrap). Two important observations emerged from this study. First, the number of false positive matches in sequence appears to be low, i.e. 8/285, 2.8% of all positive identifications, using a 0.7 ppm mass accuracy window set by MALDI-FT-ICR. Second, the false positive matches in sequence were almost exclusively found among the peptides with very low signal-to-noise (S/N) ratios. Considering the number of proteomics studies being published presenting large lists of identified peptides it becomes increasingly necessary to develop more robust standards for peptide and protein identification as this is vital for the interpretation of published proteomics results [61]. Thus it may be prudent to add more information to the identification process, such as accurate mass of the parent ion; incorporation of sequence information using sophisticated software applications [62]; more rigorous software-based cut-off values for ion scores [63]; using a combination of multiple mass spectrometry techniques to identify specific targeted peptide masses. From these experiments we concluded that the reliability of identification by MS/MS of a specific peptide mass observed in a peptide profile is determined by the following four parameters: the S/N ratio of the parent ion, the experimental mass accuracy, the Mascot score, and the frequency of observation in the different samples measured.

QUALITATIVE PROTEOMICS

In our MALDI-TOF peptide profiling study, detailed in Chapter 4, we used a shotgun proteomics approach to analyze a large number of CSF samples (n = 164). These samples, taken from neurological patients, were classed into four groups according to the clinical diagnosis: MScl (n=44), clinically isolated syndrome of demyelination (CIS, n=40), other inflammatory neurological disease (OIND, n=26) and other neurological disease (OND, n=54). After tryptic digestion, the samples were measured by MALDI-TOF MS. Spectra were analysed using the R-project software package, in which a peak detection algorithm was developed. Subsequently the peak lists were compared based on ranked data (non-parametric).

Significant differences were observed in the comparisons of MScl vs. OND and CIS vs. OND. The comparisons of MScl vs. OIND, and CIS vs. OIND showed fewer significant differences. No significant differences were found in comparisons of MScl vs. CIS and OIND vs. OND. MScl and CIS had strikingly similar profiles, probably a reflection of common pathological mechanisms. Three differentially expressed proteins in the comparison of MScl vs. OND were identified: chromogranin A, a potential marker for neurodegeneration; and two important factors in complement mediated inflammatory reaction, clusterin (apolipoprotein J) and complement C3 [64].

QUANTITATIVE PROTEOMICS

In chapter 5 we introduce a new quantitative proteomics method and used this method to analyze 163 CSF samples to search for MScl biomarkers. Protein differences in patient comparison studies are virtually never black-and-white phenomena. Thus compounds will mostly still be present in a non-diseased state, but in changed concentrations [65-68]. While scoring presence and non-presence of peptide peaks is a method of monitoring differential expression, it is more elegant and more elucidating to monitor differential peptide expression in a quantitative way by measuring peak height.

The detection of the previously reported three peptides of chromogranin A, clusterin (apolipoprotein J) en complement C3 shows that these peptides are differentially abundant in these samples regardless of the mass spectrometry method used for detection [55]. Clusterin and complement C3 have been repeatedly reported in MScl studies as being elevated in body fluids, and both proteins are functional in inflammatory processes, which are considered pivotal for the pathogenesis of MScl [69, 70]. In addition, this study also identified a number of

additional peptides differentially abundant in MScl compared to controls, like for example peptides from the variable regions of immunoglobulin gamma and kappa. The implication of immunoglobulins in MScl pathogenesis has been reported repeatedly. Elevated immunoglobulin G (IgG) levels as a marker for disease activity [71, 72], extra oligoclonal IgG bands [73] as a potential marker for progression from CIS to MScl [74] and free immunoglobulin kappa light chains in MScl cerebrospinal fluid [75-77] have all been extensively researched.

MASS SPECTROMETRY TO DISTINGUISH MULTIPLE SCLEROSIS DISEASE TYPES

Recently a number of studies aiming at identification of components of MScl pathology have been published on CSF proteomics of MScl patients. However, these studies only examined the difference between CSF of patients with MScl and CSF of patients with inflammatory and non-inflammatory neurological diseases. In Chapter 6 we discuss the results of the study in which we examined the CSF proteomic profile of two different MScl disease types, relapsing remitting (RR) MScl and primary progressive (PP) MScl, to find proteins that differentiate between the two disease types. Although the number of differentially abundant peptide peaks in the comparison between the two MScl types is limited from a statistical point of view, some peptides of the identified differentially abundant proteins are of interest.

Protein Jagged-1, a ligand for multiple Notch receptors and involved in the mediation of Notch signalling, which influences neuronal function and development [78], is elevated in RR MScl compared to PP MScl. The gene encoding this protein has been implicated in MScl previously [79] and network studies have pinpointed the immunomodulatory properties of Jagged-1, making it a new therapeutic target for MScl and other autoimmune diseases [80]. A second differentially abundant protein of interest that was more abundant in RR MScl than in PP MScl was vitamin D-binding protein. Impaired vitamin D homeostasis has been widely implicated in MScl for some years now [81-83], and this might be related to the differential abundance of vitamin D-binding protein. Vitamin D directly and indirectly regulates the differentiation and activation of CD4+ T-lymphocytes and can prevent the development of autoimmune processes [84, 85], leading to conclusions of potential involvement in MScl. Due to the fact that the geographic incidence of MScl indicates an increase in MScl with a decrease in sunlight exposure, combined with the knowledge vitamin D is produced in the skin by solar or UV irradiation and high serum levels of 25-hydroxyvitamin D have been reported to correlate with a

reduced risk of MScl, a protective role of vitamin D has been suggested [86]. It may be that in PP MScl this neuroprotective element is, at least partially, missing, resulting in a significantly more disabling disease manifestation.

VALIDATION OF MASS SPECTROMETRY-IDENTIFIED BIOMARKERS

Mass spectrometry based biomarker discovery studies are resource intensive with regards to machine time and data analysis. Thus many studies use relatively small cohorts of well characterized patients for discovery, followed by extensive validation of biomarker candidates using other methods, such as, for example, enzyme-linked immunosorbent assays (ELISAs) [87]. Such independent validation steps are necessary to confirm the identified biomarker as a valuable marker and not as an artifact result of the used mass spectrometry method. Validation techniques like ELISA, immunohistochemistry staining or Western Blotting rely on the generation or existence of high quality antibodies against candidate biomarkers, so validation is usually time and cost-intensive. A recently developed alternative is a mass spectrometry based method termed Multiple Reaction Monitoring Initiated Detection and Sequencing (MIDAS-MRM), which allows multiplexed detection of numerous characteristic peptides, and by inclusion of isotopically labeled peptides absolute quantitation can be obtained [88, 89].

For the studies described in this thesis we only used antibody based techniques for validation. To confirm the elevated abundance of chromogranin A in MScl we used an ELISA to measure chromogranin A concentrations in over 160 CSF samples and showed that MScl samples had higher chromogranin A concentrations than non-inflammatory control samples, confirming the proteomics analysis (t-test, $p = 0.00034$). Although the neurodegenerative component of MScl pathology has remained an enigma [90], an important role of innate immune cells and signals is obvious. It is of interest that chromogranin A can induce neurotoxicity in microglial cells [91]. It will be of interest to determine whether high CSF chromogranin A levels are associated with more aggressive clinical disease types.

Additionally, immunohistochemistry was used to determine clusterin expression levels in active and inactive MScl lesions in another, unpublished, validation experiment. Active lesions were shown to contain far more clusterin than inactive lesions (Figure 1). The association of CSF clusterin (apolipoprotein J), with MScl is of particular interest, because this molecule serves as a regulator (inhibitor) of complement activity. In addition, it is considered as an important factor in neuronal cell survival [92]. There is an earlier report that on elevated CSF clusterin levels in patients with central nervous system inflammation [93].

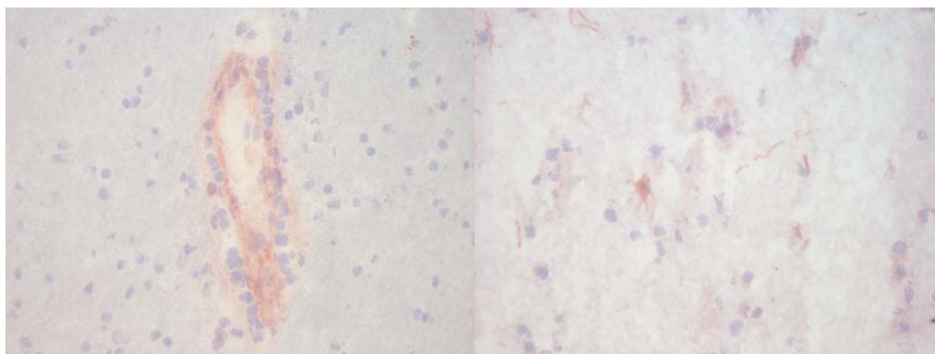


Figure 1. Immunohistochemistry staining of clusterin in MScl lesions. Active lesions (left panel) stained far more clearly for clusterin content than inactive lesions (right panel).

To validate our findings in the study comparing both MScl types (RR and PP) we performed an ELISA for vitamin D-binding protein and a Western Blot for jagged-1 protein. Validation by ELISA showed that the concentration of vitamin D-binding protein was significantly higher ($p = 0.0058$) in the RR MScl group compared to the PP MScl group. After Western Blotting, quantitative assessment of the gel bands showed that jagged-1 was indeed more abundant in RR MScl than in PP MScl ($p = 0.0124$). Separate evaluation of the levels of these two proteins does not provide high specificity or sensitivity for distinguishing between the MScl disease types, but the values increase when both proteins are evaluated together. Then the sensitivity rises to 0.8 and the specificity rises to 0.75. Although these values remain too low for truly confident separation between the disease types, they could conceivably rise further if other proteins that differentiate between MScl types are included in this analysis.

RECENT DEVELOPMENTS IN MULTIPLE SCLEROSIS CSF PROTEOMICS

In the last five years proteomics based CSF biomarker discovery studies have added a substantial number of new, potential markers for MScl. Although these studies have not resulted in new markers for the clinical diagnosis of MScl yet, the proteins published in these studies have added to the knowledge of CSF pathology in MScl. Not all these studies were focussed on detecting differences between sample groups, as some were merely aimed at detecting as many proteins as possible in MScl CSF [94, 95]. For differential peptide profiling studies several important choices need to be made; whether to pool samples or not to pool samples, which control samples need to be used, and how many samples need to be measured to obtain meaningful results. Using two-dimensional gel electrophoresis and peptide

mass fingerprinting Hammack *et al*. identified four peptides in a pooled CSF sample of three MScl patients that were not present in a pooled sample of three other patients with inflammation in the central nervous system, but no direct relation to MScl pathology was apparent [96]. Although this study identified differences between the two samples, it is important to note that these samples were pooled samples; hence a single CSF sample from one patient may skew any comparison without the researchers knowledge.

By using Surface Enhanced Laser Desorption/Ionization Time-Of-Flight (SELDI-TOF) mass spectrometry Irani *et al*. detected a cleavage product of cystatin C that was 100% specific for MScl compared to controls [97]. However, Del Boccio *et al*. later showed that this cleavage product is formed by degradation of the first eight N-terminal residues of cystatin C, and that this is a consequence of sample storage at –20°C instead of –80°C [98]. This again illustrates the importance of the sampling and sample handling phase of mass spectrometry-based proteomics. Besides controls, also Leber hereditary optic neuropathy CSF samples have been compared to MScl. The results of this study showed decreased levels of apolipoprotein A4 in MScl [99].

In another, more common type of peptide profiling experiment, using non-pooled CSF samples and a non-disease specific control group, Lehmensiek *et al*. used two-dimensional gel electrophoresis combined with MALDI-TOF mass spectrometry to identify one up-regulated and ten down-regulated proteins differing between MScl and controls and two up-regulated and eleven down-regulated proteins differing between CIS and controls. Subsequently, they used an immunoblot technique to validate their findings for apolipoprotein A1 [100]. Other, relatively small-scale, studies have added additional information on CSF proteome changes in MScl. Chiasserini *et al.* reported an increase of Ig kappa light chain in MScl compared to controls [101], and Tumani *et al.* reported on a number of proteins that are down-regulated in CIS patients that convert to RR MScl compared to CIS patients that do not progress to definitive MScl [102]. Qin *et al.* recently reported that vitamin D-binding protein levels in MScl are lower than in controls and validated those mass spectrometry findings by means of an ELISA [103], which is in agreement with our findings although in our analysis the difference between MScl and controls was not statistically significant. Analysis of naturally processed peptides bound to major histocompatibility complexes by tandem mass spectrometry was recently used to define previously known and potentially novel auto-antigens in brain autopsy material of MScl patients that are possible involved in disease induction and auto-antigen spreading [104].

Although these studies are of informative value with regards to the pathology of MScl, they are all hampered by the same issue: a low number of samples analyzed. In the peptide profiling studies described in this thesis the number of

samples analyzed is higher. Where most of the aforementioned studies analyze fewer than 30 samples from patients, we have analyzed over 160 CSF samples, increasing the statistical power of the comparisons between the sample groups. Additionally, the group of control patients was very heterogeneous, with diagnoses of the control patients ranging from migraine to Alzheimer's disease and renal carcinoma. This means that for a significant difference to be picked up in a comparison between MScl and such a control group, the difference must be far clearer and more specific than, for example, a difference between a group of MScl samples and a group of HIV samples. In a way it could thus be said, that this method would be more robust due to the diversity of the control samples. This in turn increases confidence in the identified differentially abundant proteins reported in this study, especially chromogranin A, since this protein was validated by means of an ELISA. Validation experiments were performed on both the mass spectrometry samples, and a new set of CSF samples from patients was used as an external validation sample set. The eleven published studies differentiating between MScl and various control samples, including the main findings, are listed in table 1.

Although the number of samples in the peptide profiling experiments of the RR MScl and PP MScl samples is smaller, it is important to note that this is the first in-depth survey of significant differences between MScl disease types. Both pathological and clinical biomarkers to distinguish between different MScl disease types are extremely desirable. Pathological markers could potentially elucidate the differences in clinical manifestation of the disease types, while clinical markers would enable clinicians to differentiate between disease types immediately at the time of a first CIS attack and subsequently adjust the applied treatment accordingly. The observed results could be very enlightening regarding the neuroprotective elements of vitamin D, and vitamin D-binding protein, which appears to be present in RR MScl and not in PP MScl patients.

The logical next step in these analyses is to place the differentially abundant proteins into biological context. Although proteins such as complement C3, clusterin (apolipoprotein J) and immunoglobulins are involved in the inflammatory processes of MScl, they are not specific to the disease and the situation is similar for chromogranin A and the neurodegenerative component of MScl. Also protein jagged-1 and vitamin D-binding protein may be involved in disease processes, like for example damage or damage prevention due to disease processes. However, they are not fundamental triggers of a biological cascade, but rather consequences and effects. We tried to place the differentially abundant proteins in biological processes and functions by means of network analysis. One of the available software options, as demonstrated in Chapter 6 of this thesis, is Ingenuity Pathway Analysis (Ingenuity Systems). This type of software is able to link the differentially abundant proteins

Table 1. An overview of the CSF proteomic profiling studies of MScl patients, including the number of samples measured, separation technologies used (like for example 2 dimensional gel electrophoresis (2D-GE) or isoelectric focussing (IEF)), the applied mass spectrometry and the protein database used, as well as the main results and validated differential abundant proteins between MScl and controls by independent technology.

Study	Year	Total # of samples	Pooled	Controls	Seperation technology	Mass spec-trometry	Database	Differences	Validation by independent technique
Hammack *et al.* [96]	2004	6	Yes	Viral meningitis, CNS sarcoid	2D-GE	MALDI-TOF	NCBI	tetranectin, SPARC-like protein, etc.	-
Irani *et al.* [97]	2006	133	No	OND, HIV infection, transverse myelitis	Cation exchange array	SELDI-TOF	?	cystatin C	cystatin C [a]
Noben *et al.* [95]	2006	14	Yes	Cancer	2D-LC	ESI-LCQ	UniProt and IPI	complement C3, contactin-1, etc.	-
O'Connor *et al.* [67]	2006	26	No	Normal controls	LC	ESI-TOF	?	alpha-1-antitrypsin, haptoglobin, etc.	-
Lehmensiek *et al.* [100]	2007	48	No	Headaches	2D-GE and IEF	MALDI-TOF	NCBI	Apolipoproetin E, IgK	-
D'Aguanno *et al.* [99]	2008	15	No	Healthy, Leber optic neuropathy	2D-GE	MALDI-TOF	NCBI	IgK	-
Stoop *et al.* [64]	2008	164	No	OND and OIND	Direct measurement	MALDI-TOF	UniProt	chromogranin A, clusterin, complement C3, etc.	chromo-granin A
Chiasserini *et al.* [101]	2008	31	No	Normal controls	2D-GE	ESI-LCQ	?	IgK, apolipoprotein A1, etc.	IgK light chain
Qin *et al.* [103]	2008	20	No	OND	2D-GE	MALDI-TOF	SwissProt	vitamin D-binding protein, haptoglobin, etc.	vitamin D-binding protein
Stoop *et al.* [55]	2009	163	No	OND and OIND	Direct measurement	MALDI-FT-ICR (identification by NanoLC-Orbtirap)	UniProt	chromogranin A, clusterin, complement C3, IgK, haptoglobin, etc.	-
Tumani *et al.* [102]	2009	16	No	CIS	2D-GE	MALDI-TOF	NCBI	fetuin-A, apolipoprotein A4	fetuin-A
Stoop *et al.* (Chapter 6, this thesis)	2010	31	No	Neurologically normal controls	Direct measurement	MALDI-FT-ICR (identification by NanoLC-Orbtirap)	UniProt	jagged-1, vitamin D binding protein [b]	jagged-1, vitamin D binding protein [b]

[a] Later shown by Del Boccio *et al.* to be a freezing artefact [98].
[b] Comparison between PP MScl and RR MScl.

to biological networks and pathways. However, for a more in-depth understanding of the role of these proteins in MScl disease processes, each protein must be individually assessed to determine the relevance in MScl. Although this may not result in a marker for clinical diagnosis, it will facilitate a deeper understanding of the role of these proteins in MScl pathogenesis.

FUTURE PROSPECTS

The mass spectrometry field is driven by a rapid succession of newly developed techniques and machines, which open up new options for proteomics, genomics, metabolomics and transcriptomics research [105, 106]. One of the key examples of this is the 2005 introduction of Orbitrap mass spectrometry to the field [107]. This machine performs mass analysis by radially trapping ions about a central spindle electrode. Due to its high resolution, high mass accuracy and good dynamic range the Orbitrap proved to be very suitable for proteomics research. Another interesting development has been the introduction of targeted quantitative proteomics analysis by MALDI triple quadrupole (2007), using the multiple reaction monitoring (MRM) technique, which allows for very accurate quantitation of selected compounds in complex bio-fluids [108]. Both methods offer interesting prospects for further analysis of CSF from MScl patients. Validation of differentially abundant proteins identified by mass spectrometry remains an issue because ELISA kits and antibodies are simply not available for all proteins. MRM experiments on a triple quadrupole mass spectrometer using either a MALDI or an ESI source are interesting alternatives to the antibody-based validation techniques, as it has been shown that triple quadrupole mass spectrometry can be used of quantitation of peptides [109]. However, apart from the pursuit of technological progress in protocols and instruments, stringent comparative analyses of different approaches is critical for the full development of mass spectrometry-based proteomics, because even though its potential to advance biology is nearly unlimited, so is its potential for generating poor quality data [110].

In the peptide profiling experiments described in this thesis, a MALDI mass spectrometry method without prior fractionation of the samples has been used for the profiling step of the experiments. Although this theoretically allows for an analysis of all the components of the sample, the reality is that the low abundant peptides cannot be identified in this way due to masking by ion suppression and the fact that the abundances are simply too low for detection. By adding a separation or enrichment step prior to mass spectrometry to the method, employing for example a nanoLC-ESI-Orbitrap strategy, the masking effect of the high abundant

proteins is reduced and low abundant proteins can be detected. Until very recently it was the analysis of such data that proved difficult, if not impossible. As the separation step (nano-scale liquid chromatography) is performed for all the samples, the peptides elute with a retention time that is similar, but not identical, for each sample. Consequently, identical peptides with small differences in retention time must be aligned in order to be able to make differential comparisons between the groups possible. Only very recently commercial software packages, for example Progenesis LC-MS (Nonlinear Dynamics) and Sieve (Thermo Fisher Scientific), have been introduced that are able to do this in a high through-put, and reliable way, allowing for peptide profiling experiments that delve deeper into the proteome than previously was possible. Peptide profiling of MScl and control CSF samples using this approach may yield additional, more specific MScl CSF biomarkers. However, this would require reproducible enrichment or fractionation methods, which, because of the low sample volume and low protein concentration of CSF, requires extensive method development and optimization.

Animal models for MScl, like experimental autoimmune encephalomyelitis (EAE), are of great interest for proteomics based studies. It would be extremely interesting to compare these animal models to a human MScl CSF samples on a proteomics and metabolomics level, as this could give valuable insights in the usefulness of CSF of animal EAE as a model for MScl. Furthermore, although it is difficult, due to CSF sampling regulations, to obtain CSF samples for longitudinal studies it would be very interesting to examine the effect of drug treatment on the CSF proteome of MScl patients. A difference between responders and non-responders to treatment in MScl has recently been identified on the genomics level [111]. It would be interesting to evaluate the difference between responders and non-responders also on a CSF proteomics level. Finally, it would be interesting to expand the number of samples used in the peptide profiling study of the two different MScl types. There is a dearth of literature relating to differences on a CSF proteome level between different MScl disease types. Although the disease courses of different MScl types are clinically quite distinct, it is still unclear whether PP MScl and RR MScl are two separate diseases or two different clinical subtypes of MScl. CSF proteome studies can be helpful in elucidating differences and similarities between these MScl disease types. Proper biobanking of CSF samples from MScl patients to acquire enough statistical power to execute such proteomics experiments is essential for success. Therefore worldwide initiatives in biobanking are indispensable to reach the goal of validating candidate biomarkers in larger cohorts of MScl patients.

There has been great progress in the field of tandem mass spectrometry proteomics informatics in the last five years. However, there does remain a barrier

for many research groups to use non-commercial tools, often due to a lack of user-friendliness of these tools [112]. Commercial tools such as Mascot and SEQUEST are more readily used, yet the lack of a single analysis method employed by all researchers makes evaluation and reproduction of research results extremely difficult. Many research groups have created their own protein databases and use homemade analysis tools. This is not a negative development in itself, since the availability of open-source search algorithms, like for example X!Tandem, makes it feasible to further improve peptide identification on the basis of consensus scoring [113]. But it does make comparison of different studies complex and sometimes almost impossible. The detection of a differentially abundant protein for a specific disease using a specific database search engine and a specific database is difficult to compare to other results acquired by different means. Additionally, even though the raw mass spectrometry results on identical samples measured in different laboratories appear to be reproducible, it is the analysis of the large amount of data generated in proteomics projects that is currently the major bottleneck [114]. Especially the area of search engines and databases needs improvement to bolster the reproducibility of proteomics-driven biomarker research [115, 116]. A potential source of improvement could be the increased use of sophisticated statistical models that predict the amount of certain types of observed fragment ions in peptide algorithms, since this information is at present underused and fully utilizing this information could improve current algorithms [117].

The work presented in this thesis details a successful application of modern mass spectrometry based proteomics methods to the field of biomarker discovery for MScl in CSF. Although the development of a new, clinical biomarker remains an elusive prospect, the identified differentially abundant proteins were shown to be of substantial interest in the context of the pathology of the disease. This included, for the first time, potential pathological biomarkers identified by proteomics techniques that differentiate between different MScl disease types.

REFERENCES

1. Bielekova, B., et al., *Development of biomarkers in multiple sclerosis.* Brain, 2004. **127**: p. 1463-78.
2. Compston, A., et al., *Multiple sclerosis.* Lancet, 2002. **359**: p. 1221-31.
3. Bakshi, R., et al., *MRI in multiple sclerosis: current status and future prospects.* Lancet Neurol, 2008. **7**: p. 615-25.
4. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2005 revisions to the "Mc-Donald Criteria".* Ann Neurol, 2005. **58**: p. 840-6.
5. Miller, D.H., et al., *Magnetic resonance studies of abnormalities in the normal appearing white matter and grey matter in multiple sclerosis.* J Neurol, 2003. **250**: p. 1407-19.
6. Neema, M., et al., *MRI in multiple sclerosis: what's inside the toolbox?* Neurotherapeutics, 2007. **4**: p. 602-17.
7. Pirko, I., et al., *Gray matter involvement in multiple sclerosis.* Neurology, 2007. **68**: p. 634-42.
8. Dousset, V., et al., *MR imaging of relapsing multiple sclerosis patients using ultra-small-particle iron oxide and compared with gadolinium.* AJNR Am J Neuroradiol, 2006. **27**: p. 1000-5.
9. Kohler, S.J., et al., *In vivo 13 carbon metabolic imaging at 3T with hyperpolarized 13C-1-pyruvate.* Magn Reson Med, 2007. **58**: p. 65-9.
10. Vellinga, M.M., et al., *Pluriformity of inflammation in multiple sclerosis shown by ultra-small iron oxide particle enhancement.* Brain, 2008. **131**: p. 800-7.
11. Friese, M.A., et al., *Autoreactive CD8+ T cells in multiple sclerosis: a new target for therapy?* Brain, 2005. **128**: p. 1747-63.
12. Raine, C.S., *Multiple sclerosis: immune system molecule expression in the central nervous system.* J Neuropathol Exp Neurol, 1994. **53**: p. 328-37.
13. Sospedra, M., et al., *Immunology of multiple sclerosis.* Annu Rev Immunol, 2005. **23**: p. 683-747.
14. Boven, L.A., et al., *Myelin-laden macrophages are anti-inflammatory, consistent with foam cells in multiple sclerosis.* Brain, 2006. **129**: p. 517-26.
15. Ramagopalan, S.V., et al., *The genetics of clinical outcome in multiple sclerosis.* J Neuroimmunol, 2008. **201-202**: p. 183-99.
16. Hensiek, A.E., et al., *Familial effects on the clinical course of multiple sclerosis.* Neurology, 2007. **68**: p. 376-83.
17. Aulchenko, Y.S., et al., *Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis.* Nat Genet, 2008. **40**: p. 1402-3.
18. Peltonen, L., *Old suspects found guilty--the first genome profile of multiple sclerosis.* N Engl J Med, 2007. **357**: p. 927-9.
19. Freedman, M.S., et al., *Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement.* Arch Neurol, 2005. **62**: p. 865-70.
20. Giovannoni, G., *Multiple sclerosis cerebrospinal fluid biomarkers.* Dis Markers, 2006. **22**: p. 187-96.
21. Giovannoni, G., et al., *Correlating immunological and magnetic resonance imaging markers of disease activity in multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1998. **64 Suppl 1**: p. S31-6.

22. Gironi, M., et al., *Immunological markers in multiple sclerosis.* Neurol Sci, 2000. **21**: p. S871-5.
23. Uccelli, A., et al., *Biological markers of the inflammatory phase of multiple sclerosis.* Neurol Sci, 2003. **24 Suppl 5**: p. S271-4.
24. Thompson, A.J., et al., *Primary progressive multiple sclerosis.* Brain, 1997. **120 (Pt 6)**: p. 1085-96.
25. Weiner, L.P., et al., *Immunological aspects of secondary progressive multiple sclerosis.* Mult Scler, 2002. **8**: p. 83-4.
26. Hauser, S.L., et al., *Cytokine accumulations in CSF of multiple sclerosis patients: frequent detection of interleukin-1 and tumor necrosis factor but not interleukin-6.* Neurology, 1990. **40**: p. 1735-9.
27. Trotter, J.L., et al., *Serum cytokine levels in chronic progressive multiple sclerosis: interleukin-2 levels parallel tumor necrosis factor-alpha levels.* J Neuroimmunol, 1991. **33**: p. 29-36.
28. Cohen, S.R., et al., *A diagnostic index of active demyelination: myelin basic protein in cerebrospinal fluid.* Ann Neurol, 1980. **8**: p. 25-31.
29. Semra, Y.K., et al., *Heightened intrathecal release of axonal cytoskeletal proteins in multiple sclerosis is associated with progressive disease and clinical disability.* J Neuroimmunol, 2002. **122**: p. 132-9.
30. Lycke, J.N., et al., *Neurofilament protein in cerebrospinal fluid: a potential marker of activity in multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1998. **64**: p. 402-4.
31. Kapaki, E., et al., *Increased cerebrospinal fluid tau protein in multiple sclerosis.* Eur Neurol, 2000. **43**: p. 228-32.
32. Colucci, M., et al., *The 14-3-3 protein in multiple sclerosis: a marker of disease severity.* Mult Scler, 2004. **10**: p. 477-81.
33. Jimenez-Jimenez, F.J., et al., *Tau protein concentrations in cerebrospinal fluid of patients with multiple sclerosis.* Acta Neurol Scand, 2002. **106**: p. 351-4.
34. Teunissen, C.E., et al., *Combination of CSF N-acetylaspartate and neurofilaments in multiple sclerosis.* Neurology, 2009. **72**: p. 1322-9.
35. Romeo, M.J., et al., *CSF proteome: a protein repository for potential biomarker identification.* Expert Rev Proteomics, 2005. **2**: p. 57-70.
36. Yuan, X., et al., *Proteomics analysis of prefractionated human lumbar cerebrospinal fluid.* Proteomics, 2005. **5**: p. 541-50.
37. Straus, S.E., et al., *How do I perform a lumbar puncture and analyze the results to diagnose bacterial meningitis?* Jama, 2006. **296**: p. 2012-22.
38. Pan, S., et al., *A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry.* Proteomics, 2007. **7**: p. 469-73.
39. Maurer, M.H., *Proteomics of brain extracellular fluid (ECF) and cerebrospinal fluid (CSF).* Mass Spectrom Rev, 2008.
40. Apweiler, R., et al., *Approaching clinical proteomics: current state and future fields of application in fluid proteomics.* Clin Chem Lab Med, 2009. **47**: p. 724-44.
41. Lippi, G., et al., *Reliability of the thrombin-generation assay in frozen-thawed platelet-rich plasma.* Clin Chem, 2006. **52**: p. 1827-8.
42. Conrads, T.P., et al., *Sampling and analytical strategies for biomarker discovery using mass spectrometry.* Biotechniques, 2006. **40**: p. 799-805.

43. Luque-Garcia, J.L., et al., *Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry*. J Chromatogr A, 2007. **1153**: p. 259-76.
44. Jiang, X., et al., *Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis.* Proteomics, 2008. **8**: p. 686-705.
45. Azarkan, M., et al., *Affinity chromatography: a useful tool in proteomics studies.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**: p. 81-90.
46. Lescuyer, P., et al., *Comprehensive proteome analysis by chromatographic protein prefractionation.* Electrophoresis, 2004. **25**: p. 1125-35.
47. McComb, M.E., et al., *Evaluation of an on-target sample preparation system for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in conjunction with normal-flow peptide high-performance liquid chromatography for peptide mass fingerprint analyses.* Rapid Commun Mass Spectrom, 2007. **21**: p. 44-58.
48. Bateman, R.J., et al., *Fluctuations of CSF amyloid-beta levels: implications for a diagnostic and therapeutic biomarker.* Neurology, 2007. **68**: p. 666-9.
49. Lefranc, D., et al., *Relevance of the quantification of apolipoprotein E in the cerebrospinal fluid in Alzheimer's disease.* Neurosci Lett, 1996. **212**: p. 91-4.
50. Becker, S., et al., *Surfaced-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer.* Ann Surg Oncol, 2004. **11**: p. 907-14.
51. Karp, N.A., et al., *Design and analysis issues in quantitative proteomics studies.* Proteomics, 2007. **7 Suppl 1**: p. 42-50.
52. Tolson, J., et al., *Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients.* Lab Invest, 2004. **84**: p. 845-56.
53. Cairns, D.A., et al., *Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison.* Proteomics, 2009. **9**: p. 74-86.
54. Dekker, L.J., et al., *A new method to analyze matrix-assisted laser desorption/ionization time-of-flight peptide profiling mass spectra.* Rapid Commun Mass Spectrom, 2005. **19**: p. 865-70.
55. Stoop, M.P., et al., *Quantitative matrix-assisted laser desorption ionization-fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple-sclerosis-related proteins.* J Proteome Res, 2009. **8**: p. 1404-14.
56. van Kampen, J.J., et al., *Quantitative analysis of antiretroviral drugs in lysates of peripheral blood mononuclear cells using MALDI-triple quadrupole mass spectrometry.* Anal Chem, 2008. **80**: p. 4969-75.
57. Ahmed, F.E., *Utility of mass spectrometry for proteome analysis: part II. Ion-activation methods, statistics, bioinformatics and annotation.* Expert Rev Proteomics, 2009. **6**: p. 171-97.
58. Theodorescu, D., et al., *Mass spectrometry based proteomics in urine biomarker discovery.* World J Urol, 2007. **25**: p. 435-43.
59. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**: p. 3551-67.
60. Yates, J.R., 3rd, et al., *Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database.* Anal Chem, 1995. **67**: p. 1426-36.
61. Tabb, D.L., *What's driving false discovery rates?* J Proteome Res, 2008. **7**: p. 45-6.
62. Khatun, J., et al., *Incorporating sequence information into the scoring function: a hidden Markov model for improved peptide identification.* Bioinformatics, 2008. **24**: p. 674-81.

63. Vasilescu, J., et al., *Systematic determination of ion score cutoffs based on calculated false positive rates: application for identifying ubiquitinated proteins by tandem mass spectrometry.* J Mass Spectrom, 2008. **43**: p. 296-304.
64. Stoop, M.P., et al., *Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry.* Proteomics, 2008. **8**: p. 1576-85.
65. Armah, H.B., et al., *Cerebrospinal fluid and serum biomarkers of cerebral malaria mortality in Ghanaian children.* Malar J, 2007. **6**: p. 147.
66. Kalinowska, A., et al., *PECAM-1, a key player in neuroinflammation.* Eur J Neurol, 2006. **13**: p. 1284-90.
67. O'Connor, K.C., et al., *Comprehensive phenotyping in multiple sclerosis: discovery based proteomics and the current understanding of putative biomarkers.* Dis Markers, 2006. **22**: p. 213-25.
68. Petzold, A., et al., *A systematic review and meta-analysis of CSF neurofilament protein levels as biomarkers in dementia.* Neurodegener Dis, 2007. **4**: p. 185-94.
69. Mar, P., et al., *A longitudinal study of immunological parameters in multiple sclerosis: cell-mediated immunity and complement profiles.* Clin Exp Immunol, 1979. **36**: p. 442-8.
70. Williams, K.C., et al., *Immunology of multiple sclerosis.* Clin Neurosci, 1994. **2**: p. 229-45.
71. Caroscio, J.T., et al., *Quantitative cerebrospinal fluid IgG measurements as a marker of disease activity in multiple sclerosis.* Arch Neurol, 1986. **43**: p. 1129-31.
72. Mehta, P.D., et al., *Quantitation of IgG and albumin in CSF and serum from multiple sclerosis patients by enzyme-linked immunosorbent assay.* Acta Neurol Scand, 1987. **75**: p. 125-9.
73. Zeman, A.Z., et al., *A study of oligoclonal band negative multiple sclerosis.* J Neurol Neurosurg Psychiatry, 1996. **60**: p. 27-30.
74. Lim, E.T., et al., *Cerebrospinal fluid levels of brain specific proteins in optic neuritis.* Mult Scler, 2004. **10**: p. 261-5.
75. Rinker, J.R., 2nd, et al., *Elevated CSF free kappa light chains correlate with disability prognosis in multiple sclerosis.* Neurology, 2006. **67**: p. 1288-90.
76. Fagnart, O.C., et al., *Free kappa and lambda light chain levels in the cerebrospinal fluid of patients with multiple sclerosis and other neurological diseases.* J Neuroimmunol, 1988. **19**: p. 119-32.
77. Rudick, R.A., et al., *Free kappa light chains in multiple sclerosis spinal fluid.* Ann Neurol, 1986. **20**: p. 63-9.
78. Gaiano, N., et al., *The role of notch in promoting glial and neural stem cell fates.* Annu Rev Neurosci, 2002. **25**: p. 471-90.
79. The Games Collaborative, G., et al., *Linkage disequilibrium screening for multiple sclerosis implicates JAG1 and POU2AF1 as susceptibility genes in Europeans.* J Neuroimmunol, 2006. **179**: p. 108-16.
80. Palacios, R., et al., *A network analysis of the human T-cell activation gene network identifies JAGGED1 as a therapeutic target for autoimmune diseases.* PLoS ONE, 2007. **2**: p. e1222.
81. Cantorna, M.T., *Vitamin D and autoimmunity: is vitamin D status an environmental factor affecting autoimmune disease prevalence?* Proc Soc Exp Biol Med, 2000. **223**: p. 230-3.
82. Lips, P., *Which circulating level of 25-hydroxyvitamin D is appropriate?* J Steroid Biochem Mol Biol, 2004. **89-90**: p. 611-4.
83. Szodoray, P., et al., *The complex role of vitamin D in autoimmune diseases.* Scand J Immunol, 2008. **68**: p. 261-9.

84. Adorini, L., *Selective immunointervention in autoimmune diseases: lessons from multiple sclerosis.* J Chemother, 2001. **13**: p. 219-34.
85. Cantorna, M.T., et al., *Mounting evidence for vitamin D as an environmental factor affecting autoimmune disease prevalence.* Exp Biol Med (Maywood), 2004. **229**: p. 1136-42.
86. Raghuwanshi, A., et al., *Vitamin D and multiple sclerosis.* J Cell Biochem, 2008. **105**: p. 338-43.
87. Latterich, M., et al., *Proteomics: new technologies and clinical applications.* Eur J Cancer, 2008. **44**: p. 2737-41.
88. Griffiths, J.R., et al., *The application of a hypothesis-driven strategy to the sensitive detection and location of acetylated lysine residues.* J Am Soc Mass Spectrom, 2007. **18**: p. 1423-8.
89. Mollah, S., et al., *Targeted mass spectrometric strategy for global mapping of ubiquitination on proteins.* Rapid Commun Mass Spectrom, 2007. **21**: p. 3357-64.
90. Zipp, F., et al., *The brain as a target of inflammation: common pathways link inflammatory and neurodegenerative diseases.* Trends Neurosci, 2006. **29**: p. 518-27.
91. Taylor, D.L., et al., *Activation of group II metabotropic glutamate receptors underlies microglial reactivity and neurotoxicity following stimulation with chromogranin A, a peptide up-regulated in Alzheimer's disease.* J Neurochem, 2002. **82**: p. 1179-91.
92. Li, W., et al., *Prenatal exposure of rats to Ginkgo biloba extract (EGb 761) increases neuronal survival/growth and alters gene expression in the developing fetal hippocampus.* Brain Res Dev Brain Res, 2003. **144**: p. 169-80.
93. Polihronis, M., et al., *Elevation of human cerebrospinal fluid clusterin concentration is associated with acute neuropathology.* J Neurol Sci, 1993. **115**: p. 230-3.
94. Dumont, D., et al., *Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients.* Proteomics, 2004. **4**: p. 2117-24.
95. Noben, J.P., et al., *Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography, and mass spectrometry.* J Proteome Res, 2006. **5**: p. 1647-57.
96. Hammack, B.N., et al., *Proteomic analysis of multiple sclerosis cerebrospinal fluid.* Mult Scler, 2004. **10**: p. 245-60.
97. Irani, D.N., et al., *Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis.* Ann Neurol, 2006. **59**: p. 237-47.
98. Del Boccio, P., et al., *Cleavage of cystatin C is not associated with multiple sclerosis.* Ann Neurol, 2007. **62**: p. 201-4; discussion 205.
99. D'Aguanno, S., et al., *Differential cerebro spinal fluid proteome investigation of Leber hereditary optic neuropathy (LHON) and multiple sclerosis.* J Neuroimmunol, 2008. **193**: p. 156-60.
100. Lehmensiek, V., et al., *Cerebrospinal fluid proteome profile in multiple sclerosis.* Mult Scler, 2007. **13**: p. 840-9.
101. Chiasserini, D., et al., *CSF proteome analysis in multiple sclerosis patients by two-dimensional electrophoresis.* Eur J Neurol, 2008. **15**: p. 998-1001.
102. Tumani, H., et al., *Cerebrospinal fluid biomarkers in multiple sclerosis.* Neurobiol Dis, 2009.
103. Qin, Z., et al., *Alteration of DBP levels in CSF of patients with MS by proteomics analysis.* Cell Mol Neurobiol, 2009. **29**: p. 203-10.

104. Fissolo, N., et al., *Naturally presented peptides on major histocompatibility complex I and II molecules eluted from central nervous system of multiple sclerosis patients.* Mol Cell Proteomics, 2009. **8**: p. 2090-101.
105. Gelpi, E., *From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II 1985-2000.* J Mass Spectrom, 2009.
106. Feng, X., et al., *Mass spectrometry in systems biology: an overview.* Mass Spectrom Rev, 2008. **27**: p. 635-60.
107. Hu, Q., et al., *The Orbitrap: a new mass spectrometer.* J Mass Spectrom, 2005. **40**: p. 430-43.
108. Domon, B., et al., *Challenges and opportunities in proteomics data analysis.* Mol Cell Proteomics, 2006. **5**: p. 1921-6.
109. Melanson, J.E., et al., *High-coverage quantitative proteomics using amine-specific isotopic labeling.* Proteomics, 2006. **6**: p. 4466-74.
110. Mann, M., *Comparative analysis to guide quality improvements in proteomics.* Nature Methods, 2009. **6**: p. 717-719.
111. Comabella, M., et al., *Genome-wide scan of 500,000 single-nucleotide polymorphisms among responders and nonresponders to interferon beta therapy in multiple sclerosis.* Arch Neurol, 2009. **66**: p. 972-8.
112. Deutsch, E.W., et al., *Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics.* Physiol Genomics, 2008. **33**: p. 18-25.
113. Kapp, E.A., et al., *An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis.* Proteomics, 2005. **5**: p. 3475-90.
114. Nesvizhskii, A.I., et al., *Analysis and validation of proteomic data generated by tandem mass spectrometry.* Nat Methods, 2007. **4**: p. 787-97.
115. Bell, A.W., et al., *A HUPO test sample study reveals common problems in mass spectrometry-based proteomics.* Nat Methods, 2009.
116. Aebersold, R., *A stress test for mass spectrometry-based proteomics.* Nat Methods, 2009.
117. Barton, S.J., et al., *Review of factors that influence the abundance of ions produced in a tandem mass spectrometer and statistical methods for discovering these factors.* Mass Spectrom Rev, 2009. **28**: p. 177-87.

Summary

In this thesis we present the results of our cerebrospinal fluid (CSF) biomarker research by means of proteomics methods for multiple sclerosis (MScl), which is a highly heterogeneous disease of the central nervous system. The pathology of this disease is characterized by a combination of factors such as inflammation, demyelination and axonal damage. CSF is a relatively interesting body fluid in which to search for biomarkers and disease-associated proteins and peptides for MScl, due to its close proximity to disease processes. Although MScl has been extensively researched, the cause of the disease remains elusive. Advanced unbiased discovery techniques like proteomics are able to measure large numbers of samples in relatively short periods of time. These types of experiments yield large amounts of data which, when statistically analyzed, could play a significant role in the detection of novel candidate biomarkers. CSF proteomics is a rapidly developing area of MS biomarker research that, although still relatively new, has shown some interesting discoveries.

In chapter 1 the current understanding of MScl disease pathology and the status of CSF biomarker research for MScl are reviewed. Traditionally MScl has been considered an autoimmune disorder consisting of myelin autoreactive T cells that drive an inflammatory process, leading to myelin destruction. However, MScl pathology is a great deal more complicated than that, which is not surprising given the heterogeneity observed in the clinical and phenotypical features of the disease. The unique pathologic feature of MScl is the presence of multifocal demyelinated plaques scattered through the central nervous system. Biomarker discovery studies can be done in a variety of different biological matrices, like for example body fluids or tissue. In MScl, the brain lesions are of interest as they are the primary locations of disease activity. However, these can only be collected post-mortem, so for MScl biomarker studies in living subjects body fluids are a more suitable choice. CSF collection is invasive, but the close proximity to the inflammatory lesions in the central nervous system may result in a better reflection of the relevant inflammatory processes in CSF compared to, for example, blood.

In chapter 2 we describe how we used a set of CSF samples of patients without neurological disease to analyze the variation in CSF protein and metabolite abundances in a number of well-defined individual samples by multiple analytical platforms. From these experiments we concluded that the variation of protein

levels between individuals appears to be extensive as well as dependent on the proteins itself. It seems apparent that knowledge of the range of inter-individual variation of protein levels in CSF is an important instrument for assessment of potentially elevated protein levels when comparing a group of CSF samples from patients to a group of CSF controls.

Biomarker discovery studies consist of two separate steps; discovery of differentially abundant compounds, and identification of these compounds. In chapter 3 we discuss the challenges and pitfalls of the second stage of biomarker discovery; the identification of differentially abundant peptides. By measuring 15 CSF samples using two different approaches Matrix Assisted Laser Desorption/ Ionization Time-Of-Flight (MALDI-TOF) in combination with MALDI Fourier Transform Mass Spectrometry (FTMS) in one approach and Electrospray-Orbitrap (ESI-Orbitrap) in the second approach) and comparing the peptides identified by tandem mass spectrometry we were able to determine a number of factors that influence the possibility of a false match when MS/MS spectra are searched against peptide databases. So, the reliability of identification by MS/MS of a specific peptide mass observed in a peptide profile is determined by the following four parameters: the S/N ratio of the parent ion, the experimental mass accuracy, the Mascot score, and the frequency of observation in the different samples measured.

Chapter 4 details our experiments to identify CSF biomarkers for MScl by MALDI-TOF mass spectrometry using a large number of samples. A total of 164 CSF samples taken from neurological patients were classed into four groups according to the clinical diagnosis: MScl (n=44), clinically isolated syndrome of demyelination (CIS, n=40), other inflammatory neurological disease (OIND, n=26) and other neurological disease (OND, n=54). Amongst others, significant differences were observed in the comparison of MScl vs. OND. In this comparison three differentially abundant proteins were identified: chromogranin A, a potential marker for neurodegeneration; and two important factors in complement mediated inflammatory reaction, clusterin and complement C3. CSF chromogranin A levels were confirmed to be significantly elevated in the MScl group using an ELISA.

In chapter 5 a new quantitative mass spectrometry (MS) method is presented. We used this method to further identify CSF protein differences between MScl and controls, as well as to further validate previous findings. By MALDI-FTMS, 163 CSF were measured and analyzed in a quantitative manner. Subsequently the differentially expressed peptides were identified by off-line nanoLC MALDI-TOF/ TOF and on-line nanoLC ESI-Orbitrap MS. Proteins with higher expression in MScl than in non-inflammatory controls include previously reported proteins such as chromogranin A, clusterin and complement C3. In addition, we identified immunoglobulin fragments, apolipoprotein E and a substantial number of proteins that can

be easily placed in current models of MScl pathogenesis. Interestingly, most link to the innate immune arm of the disease, a process that may be directly associated with neurodegeneration. Overall, this quantitative MS study confirmed our previous findings and added a substantial amount of new, biologically interesting proteins to the ongoing efforts to identify biomarkers for MScl.

Recently a number of studies aiming at identification of components of multiple sclerosis pathology have been published on CSF proteomics of MScl patients. However, these studies only examined the difference between CSF of patients with MScl and CSF of patients with different diseases. In chapter 6 we discuss the CSF proteomic profile of two different MScl disease types, relapsing remitting (RR) MScl and primary progressive (PP) MScl, to find proteins that differentiate between the two disease types. We analyzed 10 PP MScl, 11 RR MScl and 10 non neurological control CSF samples by quantitative MALDI-FTMS mass spectrometry and identified a number of interesting proteins that were differentially abundant in the comparison of the two multiple sclerosis types. Two of these proteins were validated other methods. Protein Jagged-1 was validated by Western blot to be far more abundant in relapsing remitting multiple sclerosis than in primary progressive multiple sclerosis ($p = 0.0124$). Vitamin D binding proteins was also more abundant in relapsing remitting multiple sclerosis, which was validated by ELISA ($p = 0.0058$).

The results of our studies show that proteomics is of substantial added value to the field of CSF biomarker research for neurological disorders and that the found markers can be confidently validated by other analytical techniques. Although the presented results may not directly lead to new therapeutic targets for MScl therapy, they do add new information that will be useful for further elucidation of MScl pathology. New developments in mass spectrometry techniques and analysis software will allow for even more detailed analysis of large sample numbers and will add more differentially abundant proteins which could provide valuable clues for the ongoing efforts to better understand the disease.

Samenvatting

In dit proefschrift presenteren wij de resultaten van het uitgevoerde proteomics onderzoek in liquor cerebrospinalis van patiënten met multiple sclerose (MScl). Deze ziekte wordt gekarakteriseerd door een combinatie van factoren zoals ontsteking, demyelinatie en axonale schade, en is vaak zeer divers qua presentatie in verschillende patiënten. Hoewel er al veel onderzoek naar MScl gedaan is, is de oorzaak van de ziekte nog niet ontdekt. De ziekteprocessen van MScl zijn vooral gelocaliseerd in het centrale zenuwstelsel. Doordat het dicht bij de ziekteprocessen ligt, is liquor cerebrospinalis een interessante optie is voor onderzoek naar de ziekte. Geavanceerde analyse technieken als proteomics hebben de capaciteit om grote aantallen monsters te analyseren en leveren vaak grote hoeveelheden data op, wat na analyse een significante bijdrage kan opleveren voor de detectie van nieuwe biomarkers voor MScl.

In hoofdstuk 1 wordt de huidige kennis over de pathologie van MScl beschreven. De ziekte wordt beschouwd een auto-immuun ziekte te zijn, waarbij autoreactieve T-cellen een ontstekingsproces tegen de myeline beschermlaag van de zenuwen voortdrijven. Echter de pathologie van MScl is een stuk gecompliceerder, wat deels geïllustreerd wordt door de heterogeniteit die geobserveerd wordt in de klinische en fenotypische representatie van de ziekte. De unieke pathologische component van MScl is de aanwezigheid van de plaques op meerdere plaatsen in het centrale zenuwstelsel waar de myeline verwijderd of beschadigd is.

Een aantal liquor cerebrospinalis monsters van patiënten zonder neurologische aandoeningen zijn, in het onderzoek beschreven in hoofdstuk 2, gebruikt om de variatie in eiwit en metaboliet waarden tussen individuen te bepalen. Uit deze resultaten kon geconcludeerd worden dat de variatie in eiwit concentraties tussen individuen vrij graat is, en afhankelijk is van welk eiwit geanalyseerd wordt. Hiermee zal terdege rekening gehouden moeten worden als monsters van patiënten met een ziekte worden vergeleken met controle monsters.

Studies om biomarkers te detecteren bestaan uit twee verschillende delen; detectie van de verschillen tussen de bestudeerde monsters, en de identificatie van die verschillen. In hoofdstuk 3 worden de uitdagingen en valkuilen van het tweede gedeelte van biomarker studies, het identificeren van de verschillen tussen (groepen van) monsters, besproken. Door verschillende vormen van massaspectrometrie te combineren hebben we kunnen bepalen welke factoren van

invloed zijn op de betrouwbaarheid van de identificatie van een peptide uit een MS/MS spectrum. De volgende vier factoren bleken van belangrijke invloed te zijn; de signaalsterkte van het peptide in de massaspectrometer, de nauwkeurigheid van de massaspectrometer, de statistische score van de zoekactie in de database, en de frequentie van observatie van het peptide in de gemeten monsters.

In hoofdstuk 4 worden de experimenten beschreven die gedaan zijn om biomarkers voor MScl te identificeren door 164 liquor cerebrospinalis monsters te meten met een massaspectrometrie techniek. Significante verschillen werden onder andere gevonden als MScl monsters werden vergeleken met monsters van controle patiënten zonder ontstekingen. Drie van deze eiwitten werden geïdentificeerd, waarvan er een, chromogranine A, werd bevestigd door extra experimenten uit te voeren met een specifieke test om de concentratie van dit eiwit te bepalen.

In hoofdstuk 5 introduceren wij een nieuwe, kwantitatieve massaspectrometrie methode en wordt deze methode gebruikt om 163 liquor cerebrospinalis monsters te analyseren. De drie eiwitten die eerder gevonden werden, werden ook door deze methode als verschillen tussen MScl en controle patiënten zonder ontstekingen gevonden. Een aantal andere eiwitten, waaronder een flink aantal dat past in de huidige kennis van de pathologie van MScl, kon aan deze lijst worden toegevoegd. Het grootste gedeelte van deze eiwitten kan worden geplaatst in de ingeboren arm van het immuun systeem, een proces wat wellicht direct gekoppeld is met de neurodegeneratieve component van de ziekte.

Met uitzondering van hoofdstuk 6 van dit proefschrift worden wereldwijd eigenlijk vrijwel alle proteomics studies in liquor cerebrospinalis van MScl patiënten uitgevoerd zonder onderscheid te maken tussen de verschillende vormen van MScl. In dit hoofdstuk beschrijven wij de experimenten die gedaan zijn om verschillen te identificeren tussen twee vormen van MScl, de primair progressieve en de relapsing remitting vorm. Door middel van de kwantitatieve methode beschreven in hoofdstuk 5 werden een aantal eiwitten gevonden die verschillend waren tussen de twee vormen van MScl. Voor twee van deze eiwitten, jagged-1 en vitamine D-bindend eiwit, kan met andere validatie technieken worden bevestigd dat deze inderdaad verlaagd aanwezig waren in de primair progressive vorm van de ziekte vergeleken met de relapsing remitting vorm.

De resultaten van deze studies tonen aan dat proteomics en massaspectrometrie technieken van substanstiele toegevoegde waarde zijn in het veld van biomarker onderzoek naar neurologische ziekten en dat de door deze technieken gevonden biomarkers betrouwbaar gevalideerd kunnen worden met andere analytische technieken. Alhoewel de gepresenteerde resultaten niet direct leiden tot nieuwe therapeutische doelen voor MScl therapie wordt er wel nieuwe informatie toegevoegd aan de huidige kennis die zeer nuttig zal zijn voor verdere opheldering

van de processen die een rol spelen bij deze ziekte. Vooruitgang in massaspectrometrie technieken en analyse software zullen het in de toekomst mogelijk maken om nog meer in detail grote aantallen monsters te analyseren en meer informatie te verschaffen over de processen die een rol spelen in MScl.

Dankwoord

Het is af. Voor mijzelf ongetwijfeld de meest prettige zin om te schrijven in dit hele proefschrift. Tijdens een promotieperiode leer je jezelf behoorlijk goed kennen, en het is me vrij duidelijk geworden dat iets afmaken mij heel erg goed bevalt. Maar ik heb dit natuurlijk niet alleen gedaan. Een proefschrift schrijf je immers niet alleen. Er hebben een flink aantal mensen op verschillende manieren een bijdrage geleverd en deze mensen zou ik graag bij deze willen bedanken.

Beste Rogier, je hebt mij, een bio-farmaceut en chemicus, heel veel geleerd over klinisch onderzoek en de ziekte multiple sclerose. Ook waardeer ik je bijdrage aan de manuscripten die we samen hebben geschreven enorm, vooral omdat je door jouw klinische achtergrond altijd weer in staat was om een nieuw perspectief over de resultaten in beeld te brengen, waardoor onze werkbesprekingen en discussies altijd weer nieuwe ideeën genereerden.

Beste Theo, het is voor mij een prachtige ervaring geweest om de massaspectrometrie groep onder jouw enthousiaste leiding te zien groeien van een kleine groep met twee massaspectrometers naar ruim vijftien mensen en zeven compleet verschillende, moderne massaspectrometers. Jouw onbedwingbare optimisme is mij tijdens het promotie-onderzoek echt een enorme steun in de rug geweest en ik heb onze samenwerking als zeer prettig ervaren, en hoop dan ook dat we dit in de komende periode kunnen continueren.

Beste Peter (Sillevis Smitt), hoewel wij elkaar niet met bijzonder grote regelmaat spraken was ik elke keer weer onder de indruk van hoe goed je op de hoogte was waar ik op dat moment mee bezig was en je had ook altijd to-the-point suggesties die ik zeer waardeerde. Ook wil ik je graag bedanken dat je me de mogelijkheid hebt gegeven om op de afdeling neurologie te promoveren.

Beste Jon, heel erg bedankt voor de altijd bijzonder uitgebreide feedback die ik van je kreeg zowel bij het schrijven van dit proefschrift, als bij de keren dat we elkaar bijvoorbeeld bij de Multiple Sclerose Dagen of op de afdeling immunologie tegenkwamen.

Beste Peter (Burgers), ik denk dat er maar heel weinig mensen zijn die meer van MALDI massaspectrometrie weten. Jouw bijdrage, zowel op technisch massaspectrometrie gebied als op gebied van het opschrijven van de resultaten in wetenschappelijk engels, aan de MALDI peptide profiling artikelen heb ik heel erg gewaardeerd.

Beste Leon, zonder jou had ik nu nog met mijn handen in mijn haar gezeten over het tweede hoofdstuk van dit proefschrift. Jouw kennis op het gebied van metabolomics was onmisbaar voor het schrijven van dat manuscript.

Het in dit proefschrift beschreven werk is uitgevoerd in de proteomics vakgroep van de afdeling neurologie, waarbij er altijd hulp gevraagd kon worden van alle collega's. Beste Lennard, Jeroen, Linda, Christoph, Martijn, Lona, Henk, Dominique, Lona, Halima, Ingrid, Roland, Vaibhav, Evert-Jan en Karin, bedankt voor de prettige samenwerking en alle suggesties en hulp die ik van jullie gekregen heb.

Er gaat bij vrijwel alles wat je doet wel eens wat fout. Zo ook hier. Het gevreesde blauwe computerscherm was begin 2008 een van de naarste verassingen van mijn leven. Mijn computer deed dus letterlijk helemaal niets meer! Bedankt Eric, voor het redden van mijn harde schijf en de informatie die er allemaal opstond.

Andreas, thanks for the both amusing as well as brilliant wednesday discussions about the pitfalls of bioinformatics and related topics.

Beste Arjen, Bianca, Ronald, Esmee, Nina, Robbert en Michel, in het intensieve en drukke programma van een promotietraject zorgden jullie voor de nodige momenten van ontspanning en gezelligheid waarbij de experimenten even vergeten konden worden.

Beste Jeroen, door onze gedeelde interesse in de wetenschap zijn wij op een bepaalde manier in staat gebleken om de Noordzee te overbruggen en nog steeds elkaars klankboord te zijn. Jouw grote gedrevenheid is voor mij een prachtig, niet te evenaren, voorbeeld. Beste Coşkun, sinds het begin van mijn promotie periode zit je opgescheept met mij links van je. Onze discussies over allerlei onderwerpen heb ik altijd erg leuk gevonden. Maar ik blijf erbij: maandag is de mooiste dag van de week. Bedankt dat jullie beiden mijn paranimf wilden zijn.

Lieve An en Wim, jullie steun voor en jullie interesse in mijn studie, onderzoek en promotie heb ik altijd erg fijn gevonden. Heel erg bedankt daarvoor.

Lieve Laura, het afgelopen half jaar is ongetwijfeld het mooiste van mijn leven. Eerst met jou trouwen en daarna promoveren. Bedankt voor je onvoorwaardelijke steun, je interesse in het onderzoek en de heel erg gewaardeerde afleiding die je me tijdens het promotie traject gegeven hebt. Mahalo!

List of publications

MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, and RQ Hintzen. **Multiple sclerosis related proteins identified in CSF by advanced mass spectrometry.** *Proteomics*. (2008) 8, page 1576-1585.

MP Stoop, RJ Lamers, PC Burgers, PAE Sillevis Smitt, RQ Hintzen, and TM Luider. **The rate of false positive sequence matches of peptides profiled by MALDI MS and identified by MS/MS.** *J Proteome Res.* (2008) 7, page 4841-4847.

MP Stoop, LJ Dekker, MK Titulaer, RJ Lamers, PC Burgers, PAE Sillevis Smitt, AJ van Gool, TM Luider, and RQ Hintzen. **Quantitative matrix-assisted laser desorption ionization-fourier transform ion cyclotron resonance (MALDI-FT-ICR) peptide profiling and identification of multiple-sclerosis-related proteins.** *J Proteome Res.* (2009) 8, page 1404-1414.

MP Stoop, RQ Hintzen, and TM Luider. **Multiple sclerosis-related biomarkers** *Predictive Diagnostics and Personalized Treatment: Dream or Reality? (2009) Editor: O. Golubnitschaja*, Chapter 18, page 355-373.

T Rosenling, CL Slim, C Christin, L Coulier, S Shi, MP Stoop, J Bosman, F Suits, PL Horvatovich, N Stockhofe-Zurwieden, R Vreeken, T Hankemeier, AJ van Gool, TM Luider, and R Bischoff. **The effect of preanalytical factors on stability of the proteome and selected metabolites in cerebrospinal fluid (CSF)**. *J Proteome Res.* (2009), Epub ahead of print.

PhD portfolio

Name PhD student: Marcel P. Stoop
Erasmus MC Department: Neurology
Research School: Erasmus Postgraduate School Molecular Medicine (MolMed)
PhD period: 2004-2008
Promotor: Prof.dr. Rogier Q. Hintzen
Supervisor: Dr. Theo M. Luider

IN-DEPTH COURSES

- FTMS: Principles and Applications, 55th Conference of the American Society for Mass Spectrometry (Indianapolis, IN, USA), 2007
- Inflammatory Mechanisms in Neurodegenerative Disease, Erasmus MC, 2006
- Mass Spectrometry of Proteins and Peptides, 53rd Conference of the American Society for Mass Spectrometry (San Antonio, TX, USA), 2005
- Immunity in the Central Nervous System: MS as a Model, Erasmus MC, 2005
- Biomedical English Writing and Communication, Erasmus MC, 2005
- Biomedical Research Techniques III, Erasmus MC, 2004

PRESENTATIONS

- MP Stoop, L Coulier, T Rosenling, S Shi, AM Smolinska, L Buydens, K Ampt, C Stingl, A Dane, B Muilwijk, PAE Sillevis Smitt, RQ Hintzen, R Bischoff, SS Wijmenga, T Hankemeier, AJ van Gool, TM Luider. Proteomics and Metabolomics biomarker studies in the CSF: the first data. Oral presentation, FIGON Dutch Medicine Days, Lunteren, 2008.
- MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Identification of multiple sclerosis related proteins in cerebrospinal fluid by advanced mass spectrometry. Poster presentation. 12th Molecular Medicine Day of the Erasmus Postgraduate school Molecular Medicine, Rotterdam, 2008.

- MP Stoop, LJ Dekker, MK. Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Identification of multiple sclerosis related proteins in cerebrospinal fluid by advanced mass spectrometry. Oral presentation. 23rd congress of the European Committee for Treatment and Research in Multiple Sclerosis, Prague, Czech Republic, 2007.
- MP Stoop, LJ Dekker, PC Burgers, RQ Hintzen, TM Luider. Improving protein coverage in small volumes of cerebrospinal fluid by the combination of different types of mass spectrometers. Poster presentation, 55th Conference of the American Society for Mass Spectrometry, Indianapolis, IN, USA, 2007.
- MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, J Bosman, RA Lamers, PAE Sillevis Smitt, R. Bischoff, TM Luider, RQ Hintzen. Analysis of CSF of multiple sclerosis patients using state-of-the-art proteomics techniques. Oral presentation. MS Research Days, Hasselt, Belgium, 2007.
- MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Distinct tryptic peptide profiles in CSF of multiple sclerosis patients analysed by mass spectrometry. Oral presentation. MS Research Days, Rotterdam, 2006.
- MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Mass spectrometry analysis of cerebrospinal fluid reveals distinct patterns in MS patients. Oral presentation. 21st congress of the European Committee for Treatment and Research in Multiple Sclerosis, Thessaloniki, Greece, 2005.
- MP Stoop, LJ Dekker, MK Titulaer, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Biomarker search by high-throughput tryptic peptide profiling of cerebrospinal fluid of patients with multiple sclerosis. Poster presentation. 53rd Conference of the American Society for Mass Spectrometry, San Antonio, TX, USA, 2005.
- MP Stoop, LJ Dekker, MK Titulaer, PC Burgers, PAE Sillevis Smitt, TM Luider, RQ Hintzen. Mass spectrometry analysis of cerebrospinal fluid reveals distinct patterns in MS patients. Oral presentation, MS Research Days, Amsterdam, 2005.
- MP Stoop, LJ Dekker, JC Dalebout, MK Titulaer, RQ Hintzen, TM Luider. Searching for biomarkers for multiple sclerosis using high-throughput peptide profiling of cerebrospinal fluid. Oral presentation. Nederlandse Vereninging voor Massaspectrometrie, Haarlem, 2005.

INTERNATIONAL CONFERENCES

- 23rd congress of the European Committee for Treatment and Research in Multiple Sclerosis, Prague, Czech Republic, 2007.
- 55th Conference of the American Society for Mass Spectrometry, Indianapolis, IN, USA, 2007.
- 21st congress of the European Committee for Treatment and Research in Multiple Sclerosis, Thessaloniki, Greece, 2005.
- 53rd Conference of the American Society for Mass Spectrometry, San Antonio, TX, USA, 2005.

AWARD

- Award for best oral presentation at the 23rd congress of the European Committee for Treatment and Research in Multiple Sclerosis, Prague, Czech Republic, 2007.

Curriculum vitae

Marcel Stoop werd op 18 oktober 1977 geboren te Rotterdam. Na het behalen van zijn VWO diploma aan het Emmaus College te Rotterdam, studeerde hij Bio-Farmaceutische Wetenschappen aan de Universiteit Leiden. Tijdens deze studie liep hij stage op de afdeling Analytical Biosciences waar hij, onder begeleiding van Prof.dr. J. van der Greef en Dr. R. van der Heijden, onderzoek deed naar specifieke isolatie van fosfopeptiden door middel van chromatografie technieken. In juli 2004 behaalde hij zijn doctoraal examen van deze studie. Hierna begon hij in september 2004 aan zijn promotie onderzoek bij de Erasmus Universiteit Rotterdam, waar hij bij de afdeling Neurologie (hoofd: Prof.dr. P.A.E. Sillevis Smitt) van de Erasmus MC onder begeleiding van Prof.dr. R.Q. Hintzen en Dr. T.M. Luider heeft gewerkt aan het onderzoek dat beschreven is in dit proefschrift. Momenteel is hij op dezelfde afdeling werkzaam als post-doc onderzoeker waar hij met massaspectrometrie technieken onderzoek doet naar multiple sclerose en een diermodel voor deze ziekte, experimentele auto-immuun encefalomyelitis.