Lumbar cerebrospinal fluid proteome in multiple sclerosis: Characterization by ultrafiltration, liquid chromatography, and mass spectrometry

Peer-reviewed author version

NOBEN, Jean-Paul; DUMONT, Debora; KWASNIKOWSKA, Natalia; VERHAERT, Peter; SOMERS, Veerle; Hupperts, Raymond; STINISSEN, Piet & ROBBEN, Johan
(2006) Lumbar cerebrospinal fluid proteome in multiple sclerosis: Characterization by ultrafiltration, liquid chromatography, and mass spectrometry. In: JOURNAL OF PROTEOME RESEARCH, 5(7). p. 1647-1657.

DOI: 10.1021/pr0504788
Handle: http://hdl.handle.net/1942/1428

Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography and mass spectrometry

Jean-Paul Noben[1,5], Debora Dumont[1,5], Natalia Kwasnikowska[2], Peter Verhaert[3], Veerle Somers[1], Raymond Hupperts[4], Piet Stinissen[1] and Johan Robben[1]

[1] Hasselt University, Biomedical Research Institute and Transnationale Universiteit Limburg, School of Life Sciences, Diepenbeek, Belgium

[2] Hasselt University, Department of Mathematics-Physics-Informatics, Diepenbeek, Belgium

[3] Delft University of Technology, Faculty of Applied Sciences, Analytical Biotechnology Group, Delft, The Netherlands

[4] University Hospital Maastricht, Department of Neurology, Maastricht, The Netherlands

[5] Both authors contributed equally to this work

Corresponding author

Johan Robben

Biomedical Research Institute (BIOMED)

Hasselt University, Agoralaan building A, 3590 Diepenbeek, Belgium

E-mail: johan.robben@uhasselt.be

Fax: +32-11-26 92 09

Abbreviations: ACN; acetonitrile, CSF; cerebrospinal fluid, GPF; gas-phase fractionation, HAc; acetic acid, LC; liquid chromatography, MS; mass spectrometry,

M.S.; multiple sclerosis, MWCO; molecular weight cut-off, SCX; strong cation-exchange chromatography

Keywords: centrifugal ultrafiltration, cerebrospinal fluid, gas-phase fractionation, liquid chromatography, mass spectrometry, multiple sclerosis, validation

Abstract

Neurological diseases, including multiple sclerosis (M.S.), often provoke changes in the functioning of the endothelial and epithelial brain barriers and give rise to disease-associated alterations of the cerebrospinal fluid (CSF) proteome. In the present study, pooled and ultrafiltered CSF of M.S. and non-M.S. patients were digested with trypsin and analyzed by off-line strong cation exchange chromatography (SCX) coupled to on-line reversed-phase LC-ESI-MS/MS. In an alternative approach, the trypsin-treated sub-proteomes were analyzed directly by LC-ESI-MS/MS and gas-phase fractionation in the mass spectrometer. Taken together, both proteomic approaches in combination with a three-step evaluation process including the search engines Sequest and Mascot, and the validation software Scaffold, resulted in the identification of 148 proteins. Sixty proteins were identified in CSF for the first time by mass spectrometry.

Introduction

Multiple sclerosis (M.S.) is an autoimmune inflammatory disease of the central nervous system (CNS)[1], resulting in the aberrant targeting and destruction of the myelin sheath due to the loss of immune homeostasis [2]. The demyelination process is characterized by an inter- but not intra-individual heterogeneity, and four distinct (but overlapping) patterns of focal demyelination were identified histologically in biopsy and autopsy material [3,4]. However, disease mechanisms in M.S. remain poorly understood at the molecular level and no reliable proteinaceous disease markers are available yet [5].

The cerebrospinal fluid (CSF) fills the ventricles and external surfaces of the CNS. The circulating fluid is considered as a 'third circulation' conveying substances secreted into the CSF to many brain regions and draining waste products of cerebral metabolism [6]. CSF homeostasis depends directly on the epithelial blood-CSF barrier located at the choroid plexuses and the outer arachnoid membrane, and indirectly on the endothelial blood-brain barrier via the interstitial compartment of the brain [6,7]. Neurological diseases, including M.S. often provoke changes in the functioning of these barriers [7,8] and give rise to disease-associated alterations of the CSF proteome. Therefore, proteomic studies of CSF have been applied to study brain disorders such as Alzheimer [9-11], Creutzfeld Jacob [12,13], schizophrenia [14] and M.S. [15,16]. It is anticipated that neuroproteomics (and neurogenomics) will indicate paraclinical disease markers of diagnostic, prognostic or therapeutic value and will give insight in brain disease initiation, propagation and recovery.

The complexity of the CSF proteome can be reduced by molecular fractionation [17], a common practice in mass spectrometry-based proteomics. In this study, molecular partitioning and concomitant complexity reduction was achieved by

combining ultrafiltration (exploiting molecular size), liquid chromatography (exploiting charge and/or polarity) and intensity-based precursor ion selection in the mass spectrometer with or without additional fractionation of the gas-phase in the ion trap mass spectrometer as described hereafter.

Ultrafiltration is mainly used for solute concentration, desalting and buffer exchange. The technique, however, was also successfully adapted to a variety of biological samples for the fractionation of proteins, enabling a more thorough analysis of a low-molecular weight protein fraction depleted from otherwise interfering molecules such as serum albumin, immunoglobulins, and other abundant, high-molecular weight proteins [18-22]. Centrifugal ultrafiltration of CSF served here the same purpose with focus on proteins in the 5 to 50 kDa MW range.

Traditionally, two-dimensional gel electrophoresis (2D-GE) has been the method of choice for the ultimate separation of CSF proteins. However, electrophoretic separation and subsequent staining of lower molecular weight proteins envisaged here, are yet not that obvious. Therefore, alternative separation technologies were considered for mining the CSF proteome [17,23-31]. In the end and as a first approach, peptide-mapping by two-dimensional liquid chromatography (2D-LC) was chosen as a plausible alternative to 2D-GE, combining off-line capillary strong cation-exchange chromatography (SCX) and on-line microcapillary reversed-phase liquid chromatography coupled to electrospray ionisation tandem mass spectrometry (LC-ESI-MS/MS). In the mass spectrometer, and as a cyclic process, the three most intense precursor ions were selected automatically from an immediately preceding full scan for subsequent collision-induced fragmentation. In a second approach, ultrafiltered tryptic digests were analyzed directly by LC-ESI-MS/MS. Further reduction of sample complexity was now accomplished by gas-phase fractionation

(GPF) [32,33] in the mass spectrometer using a prescan mass width as small as 100 Th. Repetitive injections of the same sample thus resulted in a detailed analysis of precursor ions in the 400-1400 Th mass range.

Using both proteomic approaches, in total 148 non-redundant proteins were identified in pooled CSF of relapsing-remitting M.S. patients and of a 'non-M.S.' control group. Special attention was given to the evaluation and validation of the underlying mass spectrometric data. This curative process was conducted in a semi-automatic way including the search engines Sequest and Mascot, and the statistical validation software Scaffold.

Experimental sections

CSF sampling and pooling. CSF samples were collected by means of lumbar puncture, centrifuged at 250 x *g* for 10 min and the aliquoted supernatant was stored at –80°C without the addition of protease inhibitors. CSF aliquots thus obtained from eight clinically diagnosed relapsing-remitting M.S. patients (all in clinical remission state) were thawed while on ice, pooled (8 times 600 µl; 343 µg protein/ml), mixed and finally divided in two equal sub-portions (M.S.A and M.S.B, 2.4 ml each) to allow for the exploration of different analytical procedures. Detailed patient data are summarized in Table 1A. Because of bio-ethical restrictions associated with the collection of 'healthy' CSF in Belgium, six cancer patients were selected as a 'non-M.S.' control patients having normal clinical CSF parameters with no evidence for cerebral or leptomeningeal metastasis (Table 1B). The 'non M.S.' control pool was made up by combining 400 µl CSF aliquots from selected cancer patients (474 µg protein/ml).

Sample preparation. The three CSF pools were ultrafiltered using Centricon YM-50 columns (Millipore, Bedford, CA) with a nominal molecular weight cut-off (MWCO) of 50 kDa. The non-retained fraction from each pool was concentrated using Ultrafree columns (Millipore) with a nominal MWCO of 5 kDa. To the retentate (~30 µl), 70 µl denaturation solution (50mM Tris-HCl pH 8, 6M urea, 5mM dithiothreitol) was added and the sample was incubated at 56℃ for 60 min followed by the addition of 300 µl alkylating solution (33mM iodoacetic amide in 50mM NH_4HCO_3) and incubation at room temperature for 45 min. After ultrafiltration, the final 5 kDa retentate was subjected to trypsin digestion (2 µg, 37℃, overnight). Digested protein samples were taken to dryness by *in vacuo* centrifugal evaporation and the respective residues were dissolved in 25 µl solution A (0.5% (v/v) acetic acid (HAc) in water).

Two-dimensional liquid chromatography (2D-LC). Half of each protein digest was fractionated by SCX using a Hypercarb trap column (0.5 cm x 200 µm I.D.; Nanoseparations, Nieuwkoop, The Netherlands) connected by a low-dead volume T-piece to a Polysulfoethyl aspartamide column (12 cm x 200 µm I.D.; Nanoseparations). The trapped sample was transferred to the SCX column in 1 µl solution A containing 70% acetonitrile (CAN). The analytical column was eluted with a linear salt gradient (slope 15mM KCl/min) starting from 100% solution A to 100% solution A containing 250 mM KCl and 35% (v/v) ACN, to 100 % solution A containing 500 mM KCl and 35% (v/v) ACN. By adjusting the input flow rate delivered by a HP1050 pump, the trapping and analytical split flow rates were calibrated to ~5 and ~2.5 µl/min, respectively. During trapping, the flow was directed to the waste via a needle valve connected to the aforementioned T-piece. SCX fractions of subportion M.S.A were collected manually every two minutes (designated SCX_2min_M.S.A; 26

fractions, 5 µl each), while fractionation of M.S.B and the non-M.S. pool was performed each minute (designated SCX_1min_M.S.B and SCX_1min_non-M.S., respectively; 55 fractions, 2.5 µl each). SCX fraction volume was adjusted to ~25 µl with a solution containing 5% (v/v) ACN in 100 mM HAc. This solution contained 4 pg/µl of cortisone as an internal analytical standard to monitor flow stability (by retention time) and overall performance (by peak height as derived from the selected ion chromatogram for $[M+H]^+$ = 361.2 Th) of the ion trap mass spectrometer (LCQ Classic, ThermoElectron Corporation, San Jose, CA). Of each fraction, 10 µl was injected (autoinjector AS3000, ThermoElectron) and analyzed by LC-ESI-MS/MS as described previously [15], except for the gradient slope which was halved to 0.46% ACN/min.

Gas-phase fractionation (GPF) in the ion trap. From the second half of the protein digest of CSF subportion M.S.A and non-M.S., 10 µl of each was diluted seven-fold in 5% (v/v) ACN in 100 mM HAc containing internal standard. The respective dilutions (GPF_M.S.A. and GPF_non-M.S.) were analyzed directly by LC-ESI-MS/MS operated in data-dependent acquisition mode using the three most intense precursor ions in MS/MS, under the control of Xcalibur 1.3 software (ThermoElectron). For GPF, the full scan mass range in the original acquisition method (m/z 350-1500 Th) was replaced now by one of the following mass windows (one for each of the six 10 µl sample aliquots analyzed): 400-600, 600-700, 700-800, 800-900, 900-1020 and 1020-1400 Th.

Database searching, validation and reporting. Peak lists in dta file format were generated from mass spectrometric raw data files using the CreateDTA tool available in Sequest v27 within BioWorks v3.0 (ThermoElectron). Subsequently, they were handled in a stepwise manner:

Step 1. Dta files derived from each SCX and GPF fraction were examined separately with the search engine Sequest using the database UniProt/Swiss-Prot protein knowledgebase release 42.0 (subset human; 12,088 sequences) or ipi-HUMAN (EBI release 3.02.0; 49,078 sequences). Sequest parameters were set as follows: Xcorr ≥ 1.8, ≥ 2.5 or ≥ 3.5 for singly, doubly or triply charged peptide ions; delta C_n > 0.1; precursor and product ion mass tolerance ± 3 and ± 1 Da; enzyme: trypsin; one missed cleavage allowed; static chemical modification: cysteine-carbamidomethylation; dynamic chemical modification: oxidation of methionine. Sequest identifications were assembled in a DTASelect v1.9 report [34], one for each study group. Only the dta files (and the corresponding result files in out file format) appearing in these reports, were retained for further evaluation in step 2 and 3, as described below (called hereafter 'reduced data set'). In addition, the proteins (and sibling peptides) identified in both study groups were compared and differentiated with the software Contrast [34] starting from the aforesaid DTASelect report files. Contrast and DTASelect licenses were obtained from the Scripps Research Institute (La Jolla, CA).

Step 2. The reduced data set for each study group was merged and searched with Mascot (Matrix Science) against NCBInr (download date: March 25th 2005). Search parameters were: taxonomy human, precursor and product ion mass tolerance ±3 and ±1 Da, enzyme: trypsin/P, one missed cleavage allowed. This operation resulted in two DAT files, one for each study group.

Step 3. Sequest DTA/OUT files (step 1) and Mascot DAT files (step 2) were cross-validated by Scaffold v1.0.07 (Proteome Software Inc.) for the non-M.S. and the M.S. group. The peptide identification probability was set to the maximum value of 95%. Protein identification probabilities for multiple peptide assignments were set

to ≥ 98%. Because none of the single-peptide protein reached a protein identification probability higher than 93% irrespective the spectrum quality, this value was arbitrarily set to ≥ 91%.

To validate this three-step process just described, single-peptide protein identifications (so called 'one-hit wonders') were also examined with the *de novo* sequencing algorithm Lutefisk1900 (v.1.3.2.[35], database sequence option). *De novo* derived sequence candidates were hereby evaluated against their suggested peptide sequence (as returned by Sequest and Mascot) which was entered in the Lutefisk database file. Database sequences evaluated by Lutefisk as being as good as or better than the *de novo* sequences, constituted evidence for the correctness of the respective single peptide protein identifications.

The Contrast reports obtained in step 1 were manually updated with the results returned from Mascot (step 2), Scaffold (step 3) and eventually Lutefisk. A synopsis of the extensive protein inventory report (supplemental material comprising among others sequences and search engine scores) is presented in Table 3. The number of non redundant sibling peptides was derived according to the guidelines for peptides and protein identification data [36].

Two additional data analyses were performed according to the three-step evaluation process described above:

a) to calculate the false-positive error rate, both reduced data sets were analyzed using the 'sequence-reversed' databases RV_ipi_human (step 1, Sequest search) and RV_NCBInr (step 2, Mascot search). These databases were generated from the database versions described above using an in-house developed Perl script and,

b) to study the influence of fraction volume and fractionation modus on protein count the three proteomic experiments conducted in the M.S. group (SCX_1min_M.S.B, SCX_2min_M.S.A and GPF_M.S.A) were analyzed separately.

Results and Discussion

Gel-free proteomics: 2D-LC and GPF. From each of both subportions prepared from pooled relapsing-remitting M.S. CSF (M.S.A and M.S.B), one ultrafiltered, reduced, alkylated and trypsinized sample was prepared, as described under 'experimental sections'. Half of each sample was fractionated by SCX collecting 1 (M.S.B) or 2 minute fractions (M.S.A). Resulting fractions were analyzed by LC-ESI-MS/MS. A total of 17,547 and 50,745 MS/MS spectra (dta files) were recorded for SCX_2min_M.S.A and SCX_1min_M.S.B, respectively, distributed over the collected fractions as depicted in Figure 1. As an alternative to cation-exchange fractionation, the remaining of subportion M.S.A was analyzed directly by LC-ESI-MS/MS and GPF. This experiment resulted in a record of 24,608 MS/MS spectra. The other half of subportion M.S.B was left unused in this study.

As a control, pooled CSF from six non-M.S. patients was used. Half of the ultrafiltered and trypsinized sample was fractionated by SCX collecting 1 min fractions (SCX_1min_non-M.S.) and the fractions were subsequently analyzed by LC-ESI-MS/MS (Figure 1). The other half was diluted and analyzed by GPF. Taken together, this resulted in a total of 62,814 recorded MS/MS spectra for the non-M.S. group (SCX: 45,544 spectra; GPF: 17,270 spectra).

The vast amount of information coming from the assembly of these SCX and GPF experiments is a general characteristic of gel-free mass spectrometric-based proteomics, also known as the bottom-up or the shotgun proteomic approach. In contrast to 2D-GE (where one spot contained only one or a few proteins), sibling tryptic peptides probably would be dispersed over several liquid fractions (SCX) or mass windows (GPF). In these experiments, the MS workload is related to the detail of fractionation (determined by fraction volume in SCX and by mass window width in GPF).

It is obvious that both the analytical performance of the instrument and the chromatographic process need to be as stable as possible. We monitored these parameters using an analytical internal standard (see experimental section). The coefficient of variation (CV) calculated for the retention time in each of both SCX_1min experiments (n = 55 each) was ≤ 2%. Peak height CV was 26% and 36% in the M.S. and non-M.S. group, respectively.

Database searching and validation. The number of confident protein identifications derived from the MS data is on the one hand dependent on the interplay between the search engine(s), the parameter settings and the searched protein database(s). On the other hand, the validity of the outcome of a proteomic experiment is dependent on the stringency of the validation process. In this study, a three-step process was followed for the evaluation of the data. Firstly, the data sets consisting of many thousands of dta files were examined with the search engine Sequest. This operation resulted in a reduced data set freed from contaminants (e.g. keratins) and consisting of 604 and 407 dta files in the M.S. and the non-M.S. patient group, respectively. In other words, only 0.65 % of the initial data set (155,714 dta

files) was used in this study for protein identification. Secondly, the reduced data set was searched with Mascot. This engine returned a probabilistic score. Finally, the reports coming from both search engines were evaluated with Scaffold. This software probabilistically validates the peptide identifications done by Sequest and Mascot using PeptideProphet [37] and derives corresponding protein probabilities using ProteinProphet [38]. Peptides and assigned proteins meeting the preset criteria (as defined in experimental sections) were considered as confidently identified. The impact of the stepwise validation on peptide and protein counts for both groups and for each step is given in Table 2.

It should be remarked that despite the applied three-step evaluation procedure, some of the peptide assignments still remain ambiguous. These include mass spectra which, upon visual inspection, displayed a high background noise and/or showed (multiple) unassigned ion peaks. Therefore, the single-peptide protein identifications were examined additionally with Lutefisk. This showed that 90% of the sequences returned by Sequest and Mascot matched one of the *de novo* sequences proposed by Lutefisk. In the other cases (indicated in Table 3), no quality *de novo* sequences could be derived (noisy spectra) or *de novo* sequences were not in agreement with those returned by both search engines.

Finally, the false-positive error rate for the reduced data set (1011 spectra) using the 'sequence-reversed' database and following the three-step evaluation process, was 0.9%.

Study outcome. The assembly of the fractionation experiments performed on the three CSF pools followed by our stepwise validation strategy resulted in the identification of in total 148 proteins (Table 3): 80 of which were found in common for

both patient groups (Table 3a), 24 proteins were shotgunned in the non-M.S. group (Table 3b) and 44 only showed up in M.S. CSF (Table 3c). Single-peptide protein identifications amounted to 45% in both the M.S. and non-M.S. group (Table 3, Figure 3).

The non-identification of known CSF proteins in the non-M.S. groups including alpha-1-antitrypsin and serotransferrin (TRFE) does not necessarily imply that the peptides were not detected, but merely that they did not survive the stringent three step evaluation process. However, the high number of nonredundant peptides encountered in the M.S. group for TRFE on the one hand and the total lack of peptides reaching significance in the non-M.S. group remains unexplained. In two independent 1D electrophoresis experiments (of which one is presented here as supplemental material) TRFE was identified together with albumin as >50 kDa proteins that survived inexpectedly the 50 kDa ultrafiltration step, indicating an absolute MWCO being significantly higher than the specified nominal MWCO. Subsequently, we should have encountered TRFE in both study groups. Van Rensburg *et al.* [39] described a selective truncation of TRFE under conditions of inflammation which resulted in the appearance of a lower molecular weight protein designated as 'toxiferrin'. Since TRFE was detected here in M.S. patients only, conversion of TRFE to toxiferrin might have occurred. Therefore, M.S. and non-M.S. CSF samples were analyzed with 1-D electrophoresis followed by western blotting using in-house raised monoclonal antibodies. Beside TRFE (77kDa) several lower molecular weight bands were immunodetected with no substantial difference in both groups. The reason why we did not detect peptides reaching significance in the non-M.S. group therefore remains obscure and might be related to the spin column itself, but certainly is unrelated to the mass spectrometrical process since analytical quality

criteria (as judged from added internal standard) were always met in this study. Indeed, 1D-GE analysis of a CSF (5-50 kDa) ultrafiltrate evidenced two weak gel bands that were identified by mass spectrometry as serotransferrin and serum albumin (supplemental material), To the best of our knowledge, 60 proteins were detected here for the first time in CSF by mass spectrometry (Table 4), although some of these proteins were already indirectly evidenced by others in CSF (using alternative detection methods) or in human brain studies (Table 3 and 4). Five identifications remained ambiguous: oxytocin or vasopressin-neurophysin 2-copeptin (in M.S. and non-M.S. group), cytoplasmic actin 1 or 2 (M.S. group), DNA-binding protein A or B (M.S. group), metallothionein I or II (M.S. group) and ubiquitin-like protein SMT3A or SMT3B (non-M.S. group).

Since the M.S. and non-M.S. group are not age-matched, age-related changes in the protein profiles might have occurred in this study. Comparing older individuals (as is the case in the non-M.S. group) and younger individuals (as is the case in the M.S. group) it was shown that 30 of the 300 identified CSF proteins displayed a >20% fold abundance change [40]. From these 30 proteins five were also detected in our study: four proteins were present in both study groups (KLK6, CO4, PGCN and APOE) and had a comparable number of sibling peptides (see Table 3a), while one other protein (PLMN) occurred in the non-M.S. group only (see Table 3b). PLMN, however, was shown to be downregulated in the elderly patient group [40]. Based on these findings coming from a comparison of a limited set of proteins shown to be prone to age-related changes, we conclude that there is no evidence that the age difference between the M.S. and non-M.S. group had a substantial impact on the outcome of this qualitative study.

Reducing SCX fraction volume from ~5 µl to ~2.5 µl by lowering the fraction collection time from two to one minute had a beneficial effect on the number of detected proteins (Figure 3). Of the 113 proteins identified in both SCX experiments, 52 proteins were 'unique' for SCX_1min_M.S.B versus 8 proteins for SCX_2min_M.S.A This can largely be attributed to the higher number of different parent ions detected in the mass spectrometer in the data-dependent mode upon fraction volume reduction. Ion suppressive effects in the ion source might have become reduced as well. As a consequence, more MS/MS spectra were recorded upon fraction volume reduction (*vide supra*). The 'unique' proteins identified in SCX_1min_M.S.B or SCX_2min_M.S.A are indicated in Table 3.

GPF_M.S.A and SCX_2min_M.S.A experiments resulted in a comparable protein count, but SCX_1min_M.S.B outnumbered both (Figure 3). Combining SCX_1min_M.S.B and GPF_M.S.A experiments, 118 out of the 124 proteins (95%) were covered for the M.S. patient group (Figure 2). This observation rationalized to some extent the non conductance of a SCX_2min experiment in the non-M.S. group. The proteins identified in the M.S. group by GPF_M.S.A only are indicated in Table 3.

Potential CSF biomarkers in multiple sclerosis. Some of the identified proteins were indicated as M.S.-related biomarkers in a paper [5] critically reviewing published studies in the field of M.S. (period 1982-2002). Proteins reported to be implicated in M.S. pathogenesis (and encountered here only in the M.S. group) were: complement C3, peptidyl-glycine α–amidating monooxygenase and members of the heat shock protein (hsp) family hsp90, glucose-regulated protein 78 and endoplasmatic reticulum protein 29. Hsp proteins were entitled as biomarkers with potential for further research [5] since dysregulation in the hsp system was found to be

the most prominent and consistent result of gene expression studies in M.S. and other autoimmune diseases. Other proposed M.S.-related biomarkers [5] were detected in our study in both the M.S. group and non-M.S. group i.e. complement factors C4, CD59 (the regulator of complement activation and potent inhibitor of the membrane attack complex), osteopontin, amyloid A protein (here amyloid beta A4) and the neural cell adhesion molecule N-CAM (here N-CAM140). The latter was considered as a much-needed biomarker that would guide development of repair-promoting strategies in M.S. and aid in disease heterogeneity studies [5]. For the functional annotation of the tabulated proteins (Table 3) and their role in brain disease in general and M.S. in particular, we refer to Swiss-Prot and recent publications concerning these topics (Table 3).

The implication in M.S. pathogenesis of (some of) the proteins evidenced in this study and their potential as paraclinical disease markers can not be derived directly from a protein profiling study such as the one conducted here. Quantitative results will be required for those proteins (e.g. osteopontin) found in common, but perhaps significantly differing in concentration, while more experimental repetition is needed to address the reproducibility issue of the so-called 'unique' proteins found in each of both patient groups. Therefore, we evaluated one of the proteins shotgunned in the M.S. group, the proteinase inhibitor cystatin A (CYSA). Imbalance between cysteine proteinases and their inhibitors have been associated with different diseases such as rheumatoid arthritis [41], Alzheimer disease [42] and M.S. [43-44]. Although cystatin C has been studied extensively with respect to neurodegeneration [45-49], little is known about the role of CYSA. Therefore, we performed a CYSA-specific ELISA (KRKA Diagnostics, Slovenia) (Figure 4) on serum and CSF samples of M.S. an non-M.S. patients. Although two M.S. patients showed elevated CYSA levels in their CSF, the

mean CSF value was not significantly different in relapsing-remitting M.S. patients in state of remission as compared to non-M.S. patients. In serum, however, the CYSA was increased in the M.S. group (p=0.0075., unpaired one-sided t-test). Strikingly, paired data available for 1 M.S. patient showed elevated CYSA level in both serum and CSF. It should be noted that the CYSA concentration was ca. 100 and 1000 times lower than the cystatin C concentration in serum and CSF, respectively. This finding makes CYSA an additional candidate biomarker in brain disease.

Ultrafiltration. CSF used in this study was prefractionated by ultrafiltration with focus on proteins in the 5 to 50 kDa MW range. Principally, smaller (< 5kDa) and larger (> 50 kDa) proteins were discarded from this study. Therefore, ultrafiltration is a very effective method to deplete the high abundant proteins, like albumin (66 kDa), serotransferrin (77 kDa), and IgG (150 kDa) from CSF. This step, however, possibly would also eliminate < 50 kDa non-redundant proteins forming > 50 kDa homomeric or heteromeric complexes, as was evidenced indirectly by 2D-GE of ultrafiltered plasma [50]. Also in this study, CSF proteins such as apoliprotein A-I to IV, vitamin D binding protein, retinol binding protein and tetranectin were not detected, whereas they were readily identified in the proteome of untreated [15,16] or organic solvent-treated CSF [50,51]. Some of these proteins are known to form specific complexes [52] and could therefore be lost upon ultrafiltration as was confirmed here by 1D-GE analysis of the retentate (supplemental material). To deal with this drawback of ultrafiltration, future research will include the use of solvent conditions disrupting protein-protein interactions prior to ultrafiltration [22]. Nevertheless, even under the conditions used in this study, the combination of ultrafiltration and liquid chromatography with or without GPF allowed for the mass spectrometric detection of a vast number of (less-abundant) yet unreported proteins in the < 50 kDa region. It

should be stressed, however, that analytes such as cytokines and chemokines (molecular weight range ~8-30 kDa) were still undetected in our experimental setting (identification limit equivalent to ~3 femtomole bovine serum albumin; data not shown). These analytes were also missing from any other human CSF proteome study published so far.

Finally, the membranes used in ultrafiltration devices are characterized by a nominal MWCO. Their ability to retain molecules depends on the solute's molecular weight and size. Solutes with molecular weights close to the membrane cut-off are known to be only partially retained and flow-through may occur to some extent [50], as illustrated in this study for serum albumin and TRFE (*vide supra*). In a total of 148 proteins identified in this study (Table 3), 100 proteins (68%) had a precursor MW lower than 50 kDa, whereas 24 proteins (16%) were situated in the 50-90 kDa range. Twenty-four proteins (16%) even had a MW of more than 90 kDa (Table 3 and 4). Their presence in the ultrafiltrate is probably due to (un)specific protein processing. As a matter of fact, some truncated form(s) of reported > 90 kDa proteins were already annotated on 2D-gels by others in the lower (< 50 kDa) gel region, including perlecan, contactin 1, complement C3 and complement C4 [53], and neural-cadherin [54].

Conclusion

Notwithstanding the unprecedented fractionating power of 2D-GE, ultrafiltration was here combined with chromatographic fractionation of CSF tryptic peptides with or without GPF. The applied ultrafiltration technique proved to be a valuable partitioning method that can be further optimized to increase specificity. The application of 2D-LC-ESI-MS/MS and 1D-LC-ESI-MS/MS (the latter with GPF) on the ultrafiltered CSF of M.S. and non-M.S. patients resulted in the confident identification

of in total 148 proteins. Although more sample consuming, the GPF approach was shown to be a complementary approach to 2D-LC-ESI-MS/MS, allowing for a more detailed examination of the mass spectrum in time when combined with 1D-LC. The power and outcome of the described shotgun strategy is strongly dependent on the validation of mass spectrometric information. Despite the availability of good search engines and validation software, this process still is a daunting task. In addition, tools for biological interpretation of proteomic data have not yet penetrated satisfactorily the proteomic community. Relief may come from platforms for global proteome profiling and biomarker discovery [55].

Finally, this study is a gel-free approach of CSF proteomes of M.S. patients. Evidence is presented for the occurrence in CSF of proteins not yet reported in previous gel-based or gel-free CSF studies. Proteins unique to M.S. (and selected non-M.S. patients) were evidenced as well. A significant elevation of CYSA was shown in serum of M.S. patients as compared to healthy controls. A follow-up study is planned to shed light on the reproducibility of other so-called ‘unique’ proteins and on their relevance with respect to CSF and M.S. pathogenesis.

Acknowledgements. The authors thank Veronique Haesen and Erik Royackers for excellent technical assistance. We also acknowledge Dr. R. Medaer and Dr. JL. Rummens for providing the CSF samples and determining the IgG index of M.S. patients. This work is supported by grants from the transnationale Universiteit Limburg. P. Verhaert is member of the Laboratory for Molecular Cell Biology of the Flanders Interuniversity Institute for Biotechnology (VIB). D. Dumont holds a fellowship from the Belgian 'Wetenschappelijk Onderzoek Multiple Sclerosis' (WOMS) Foundation.

Supporting information available:

protein inventory

1D-GE and corresponding DTASelect report

This material is available free at http://pubs.acs.org

References

(1) Wekerle, H., Linington, C., Lassmann, H., Meyermann, R., *Trends in Neurosciences* 1986, *9*, 271-277.

(2) Robinson, W. H., Utz, P. J., Steinman, L., *Curr.Opin.Immunol.* 2003, *15*, 660-667.

(3) Lucchinetti, C., Bruck, W., Parisi, J., Scheithauer, B. et al., *Ann.Neurol.* 2000, *47*, 707-717.

(4) Kornek, B. and Lassmann, H., *Brain Res.Bull.* 2003, *61*, 321-326.

(5) Bielekova, B. and Martin, R., *Brain* 2004, *127*, 1463-1478.

(6) Segal, M. B., *Cell Mol.Neurobiol.* 2000, *20*, 183-196.

(7) Marchi, N., Cavaglia, M., Fazio, V., Bhudia, S. et al., *Clin.Chim.Acta* 2004, *342*, 1-12.

(8) Reiber, H., *Mult.Scler.* 1998, *4*, 99-107.

(9) Choe, L. H., Dutt, M. J., Relkin, N., Lee, K. H., *Electrophoresis* 2002, *23*, 2247-2251.

(10) Davidsson, P., Sjogren, M., Andreasen, N., Lindbjer, M. et al., *Mol.Brain Res.* 2002, *109*, 128-133.

(11) Puchades, M., Hansson, S. F., Nilsson, C. L., Andreasen, N. et al., *Mol.Brain Res.* 2003, *118*, 140-146.

(12) Hsich, G., Kenney, K., Gibbs, C. J., Lee, K. H. et al., *N.Engl.J.Med.* 1996, *335*, 924-930.

(13) Guillaume, E., Zimmermann, C., Burkhard, P. R., Hochstrasser, D. F. et al., *Proteomics.* 2003, *3*, 1495-1499.

(14) Harrington, M. G., Merril, C. R., Torrey, E. F., *Clin.Chem.* 1985, *31*, 722-726.

(15) Dumont, D., Noben, J. P., Raus, J., Stinissen, P. et al., *Proteomics.* 2004, *4*, 2117-2124.

(16) Hammack, B. N., Fung, K. Y., Hunsucker, S. W., Duncan, M. W. et al., *Mult.Scler.* 2004, *10*, 245-260.

(17) Stasyk, T. and Huber, L. A., *Proteomics.* 2004, *4*, 3704-3716.

(18) Wagner, K., Miliotis, T., Marko-Varga, G., Bischoff, R. et al., *Anal.Chem.* 2002, *74*, 809-820.

(19) Chernokalskaya, E., Gutierrez, S., Pitt, A. M., Leonard, J. T., *Electrophoresis* 2004, *25*, 2461-2468.

(20) Gretzer, M. B., Chan, D. W., van Rootselaar, C. L., Rosenzweig, J. M. et al., *Prostate* 2004, *60*, 325-331.

(21) Merrell, K., Southwick, K., Graves, S. W., Esplin, M. S. et al., *J.Biomol.Tech.* 2004, *15*, 238-248.

(22) Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J. et al., *Mol.Cell Proteomics* 2003, *2*, 1096-1103.

(23) Heine, G., Zucht, H. D., Schuhmann, M. U., Burger, K. et al., *J.Chromatogr.B* 2002, *782*, 353-361.

(24) Stark, M., Danielsson, O., Griffiths, W. J., Jornvall, H. et al., *J.Chromatogr.B* 2001, *754*, 357-367.

(25) Maccarrone, G., Milfay, D., Birg, I., Rosenhagen, M. et al., *Electrophoresis* 2004, *25*, 2402-2412.

(26) Wetterhall, M., Palmblad, M., Hakansson, P., Markides, K. E. et al., *J.Proteome.Res.* 2002, *1*, 361-366.

(27) Ramstrom, M., Palmblad, M., Markides, K. E., Hakansson, P. et al., *Proteomics.* 2003, *3*, 184-190.

(28) Romijn, E. P., Krijgsveld, J., Heck, A. J., *J.Chromatogr.A* 2003, *1000*, 589-608.

(29) Wang, H. and Hanash, S., *J.Chromatogr.B* 2003, *787*, 11-18.

(30) Bischoff, R. and Luider, T. M., *J.Chromatogr.B* 2004, *803*, 27-40.

(31) Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. et al., *J.Proteome.Res.* 2003, *2*, 43-50.

(32) Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H. et al., *Proteomics.* 2001, *1*, 93-107.

(33) Blonder, J., Rodriguez-Galan, M. C., Lucas, D. A., Young, H. A. et al., *Biochim.Biophys.Acta* 2004, *1698*, 87-95.

(34) Tabb, D. L., McDonald, W. H., Yates, J. R., III, *J.Proteome.Res.* 2002, *1*, 21-26.

(35) Taylor, J. A. and Johnson, R. S., *Rapid Commun.Mass Spectrom.* 1997, *11*, 1067-1075.

(36) Carr, S., Aebersold, R., Baldwin, M., Burlingame, A. et al., *Mol.Cell Proteomics.* 2004, *3*, 531-533.

(37) Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., *Anal.Chem.* 2003, *75*, 4646-4658.

(38) Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., *Anal.Chem.* 2002, *74*, 5383-5392.

(39) van Rensburg, SJ., van Zyl, J., Hon, D., Daniels, W. et al., *Metab Brain Dis*. 2004 ,19, 97-112

(40) Zhang, J., Goodlett, D. R., Peskind, E. R., Quinn, J. F. et al., *Neurobiol.Aging* 2005, *26*, 207-227.

(41) Trabandt, A., Gay, R. E., Fassbender, H. G., and Gay, S., *Arthritis Rheum.*1991,*34*, 1444-1451.

(42) Nakamura, Y., Takeda, M., Suzuki, H., Hattori, H., Tada, K., Hariguchi, S., Hashimoto, S., and Nishimura, T., *Neurosci. Lett.* 1991, *130*, 195-198.

(43) Bever, C. T., Jr. and Garver, D. W., *J. Neurol. Sci.* 1995,*131*, 71-73.

(44) Henskens, Y. M., Veerman, E. C., and Nieuw Amerongen, A. V., *Biol. Chem. Hoppe Seyler* 1996,*377*, 71-86.

(45) Pirttila, T. J., Manninen, A., Jutila, L., Nissinen, J., Kalviainen, R., Vapalahti, M., Immonen, A., Paljarvi, L., Karkola, K., Alafuzoff, I., Mervaala, E., and Pitkanen, A., *Ann. Neurol.* 2005,*58*, 211-223.

(46) Cathcart, H. M., Huang, R., Lanham, I. S., Corder, E. H., and Poduslo, S. E., *Neurology* 2005,*64*, 755-757.

(47) Olsson, T., Nygren, J., Hakansson, K., Lundblad, C., Grubb, A., Smith, M. L., and Wieloch, T., *Neuroscience* 2004,*128*, 65-71.

(48) Sastre, M., Calero, M., Pawlik, M., Mathews, P. M., Kumar, A., Danilov, V., Schmidt, S. D., Nixon, R. A., Frangione, B., and Levy, E., *Neurobiol. Aging* 2004,*25*, 1033-1043.

(49) Irani, D. N., Anderson, C., Gundry, R., Cotter, R., Moore, S., Kerr, D. A., McArthur, J. C., Sacktor, N., Pardo, C. A., Jones, M., Calabresi, P. A., and Nath, A. (2006) *Ann. Neurol. 59*, 237-247.

(50) Georgiou, H. M., Rice, G. E., Baker, M. S., *Proteomics.* 2001, *1*, 1503-1506.

(51) Hansson, S. F., Puchades, M., Blennow, K., Sjogren, M. et al., *Proteome.Sci.* 2004, *2*, 7-17.

(52) Nielsen, B. B., Kastrup, J. S., Rasmussen, H., Holtet, T. L. et al., *FEBS Lett.* 1997, *412*, 388-396.

(53) Finehout, E. J., Franck, Z., Lee, K. H., *Electrophoresis* 2004, *25*, 2564-2575.

(54) Yuan, X. and Desiderio, D. M., *Proteomics.* 2005, *5*, 541-550.

(55) Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G. et al., *Mol.Cell Proteomics.* 2004, *3*, 984-997.

(56) van Kammen, D. P., Guidotti, A., Neylan, T., Guarneri, P. et al., *Eur.Arch.Psychiatry Clin.Neurosci.* 1994, *244*, 216-222.

(57) van Kammen, D. P., Guidotti, A., Kelley, M. E., Gurklis, J. et al., *Biol.Psychiatry* 1993, *34*, 515-522.

(58) Bergquist, J., Palmblad, M., Wetterhall, M., Hakansson, P. et al., *Mass Spectrom.Rev.* 2002, *21*, 2-15.

(59) Davidsson, P., Paulson, L., Hesse, C., Blennow, K. et al., *Proteomics.* 2001, *1*, 444-452.

(60) Sickmann, A., Dormeyer, W., Wortelkamp, S., Woitalla, D. et al., *Electrophoresis* 2000, *21*, 2721-2728.

(61) Sickmann, A., Dormeyer, W., Wortelkamp, S., Woitalla, D. et al., *J.Chromatogr.B* 2002, *771*, 167-196.

(62) Yuan, X., Russell, T., Wood, G., Desiderio, D. M., *Electrophoresis* 2002, *23*, 1185-1196.

(63) Thibout, H., Martinerie, C., Creminon, C., Godeau, F. et al., *J.Clin.Endocrinol.Metab* 2003, *88*, 327-336.

(64) Lubec, G., Krapfenbauer, K., Fountoulakis, M., *Prog.Neurobiol.* 2003, *69*, 193-211.

(65) Davidsson, P. and Nilsson, C. L., *Biochim.Biophys.Acta* 1999, *1473*, 391-399.

(66) Davidsson, P., Folkesson, S., Christiansson, M., Lindbjer, M. et al., *Rapid Commun.Mass Spectrom.* 2002, *16*, 2083-2088.

(67) Fountoulakis, M., *Mass Spectrom.Rev.* 2004, *23*, 231-258.

(68) Lescuyer, P., Allard, L., Zimmermann-Ivol, C. G., Burgess, J. A. et al., *Proteomics.* 2004, *4*, 2234-2241.

(69) Bartosik-Psujek, H. and Archelos, J. J., *J.Neurol.* 2004, *251*, 414-420.

(70) Satoh, J., Yukitake, M., Kurohara, K., Takashima, H. et al., *J.Neurol.Sci.* 2003, *212*, 11-20.

(71) Van Everbroeck, B., Boons, J., Cras, P., *Clin.Neurol.Neurosurg.* 2005, *107*, 355-360.

(72) Newcombe, J., Eriksson, B., Ottervald, J., Yang, Y. et al., *J.Chromatogr.B* 2005, *815*, 191-202.

(73) Raymackers, J., Daniels, A., De Brabandere, V., Missiaen, C. et al., *Electrophoresis* 2000, *21*, 2266-2283.

(74) Vermes, I., Steur, E. N., Reutelingsperger, C., Haanen, C., *Mov Disord.* 1999, *14*, 1008-1010.

(75) Takayasu, M., Shibuya, M., Kanamori, M., Suzuki, Y. et al., *J.Neurosurg.* 1985, *63*, 417-420.

(76) Coplin, W. M., Longstreth, W. T., Jr., Lam, A. M., Chandler, W. L. et al., *Arch.Neurol.* 1999, *56*, 1348-1352.

(77) Kaukinen, L., Porkkala, H., Kaukinen, S., Pehkonen, E. et al., *Acta Anaesthesiol.Scand.* 2000, *44*, 361-368.

(78) Sherman, A. L., Tirschwell, D. L., Micklesen, P. J., Longstreth, W. T., Jr. et al., *Neurology* 2000, *54*, 889-894.

(79) Ando, K., Hirata, Y., Togashi, K., Kawakami, M. et al., *J.Cardiovasc.Pharmacol.* 1991, *17 Suppl 7*, S434-S436.

(80) Kraus, G. E., Bucholz, R. D., Yoon, K. W., Knuepfer, M. M. et al., *Surg.Neurol.* 1991, *35*, 20-29.

(81) Yamaji, T., Johshita, H., Ishibashi, M., Takaku, F. et al., *J.Clin.Endocrinol.Metab* 1990, *71*, 1611-1615.

(82) Ehrenreich, H., Lange, M., Near, K. A., Anneser, F. et al., *Res.Exp.Med.*1992, *192*, 257-268.

(83) Suzuki, H., Sato, S., Suzuki, Y., Oka, M. et al., *Ann.Med.* 1990, *22*, 233-236.

(84) Van Gool, S. W., De Meyer, G., Van, d., V, Vanmechelen, E. et al., *Neurotoxicology* 2004, *25*, 471-480.

(85) Li, C. Y., Ziesmer, S. C., Wong, Y. C., Yam, L. T., *Acta Cytol.* 1989, *33*, 667-673.

(86) Oppermann, M., Cols, N., Nyman, T., Helin, J. et al., *Eur.J.Biochem.* 2000, *267*, 4713-4719.

.

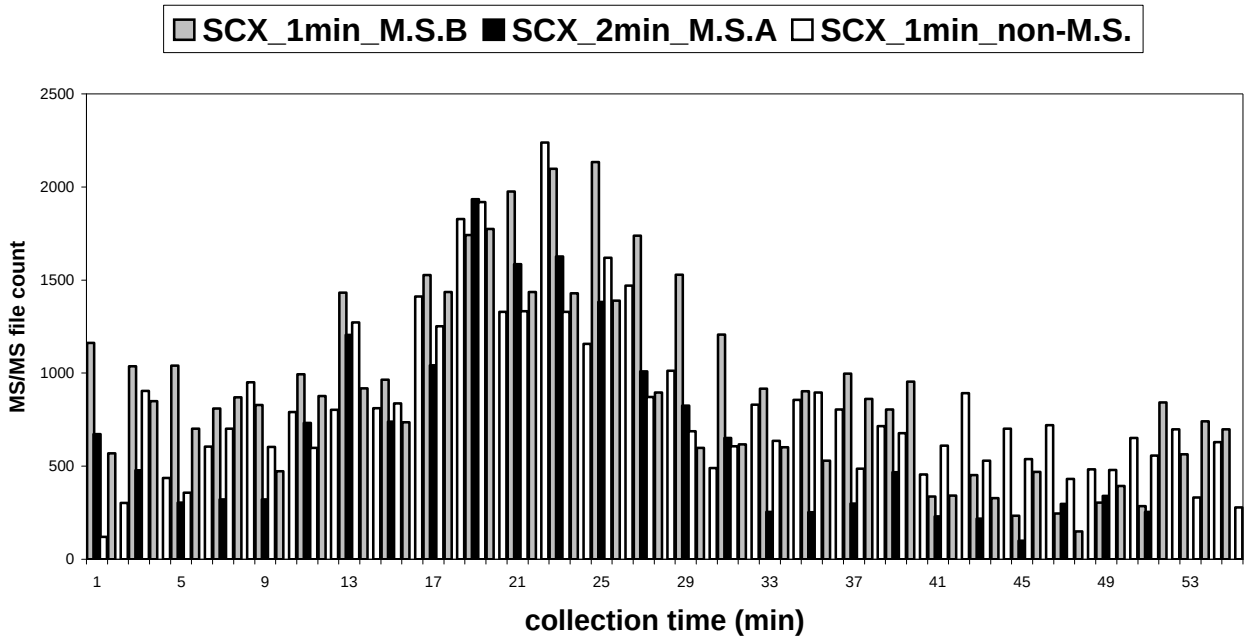

Figure 1: Number of tandem mass spectrometric files recorded at a given SCX fraction collection time

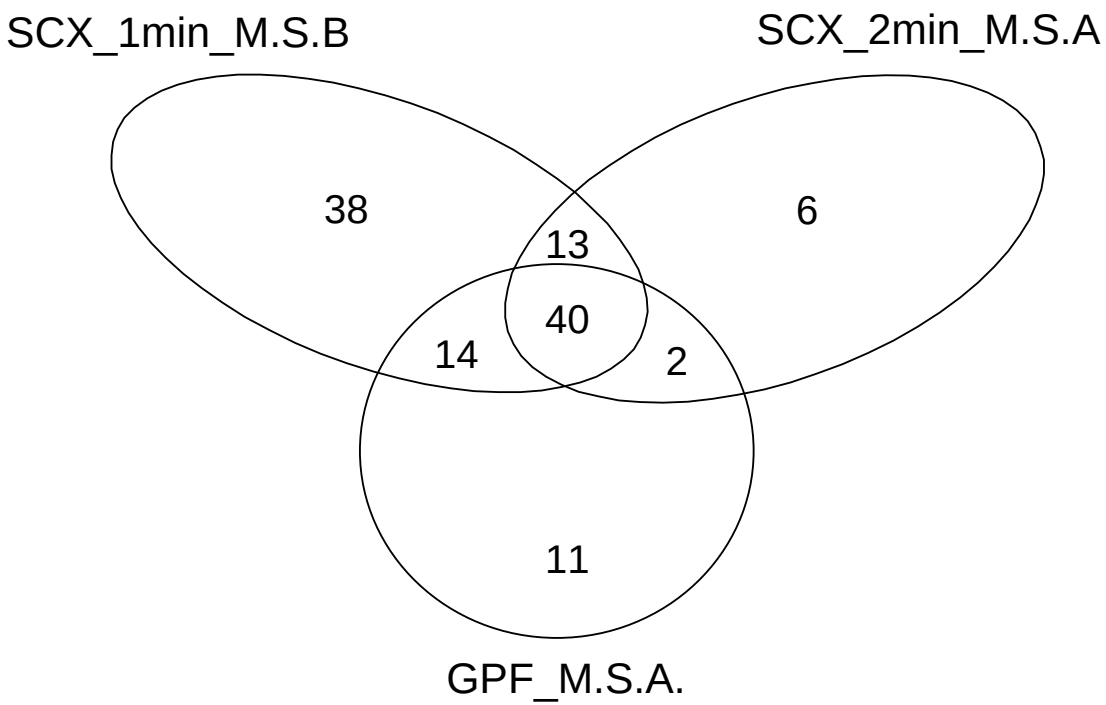

Figure 2: Number of protein identifications for each fractionation method used on CSF of M.S. patients

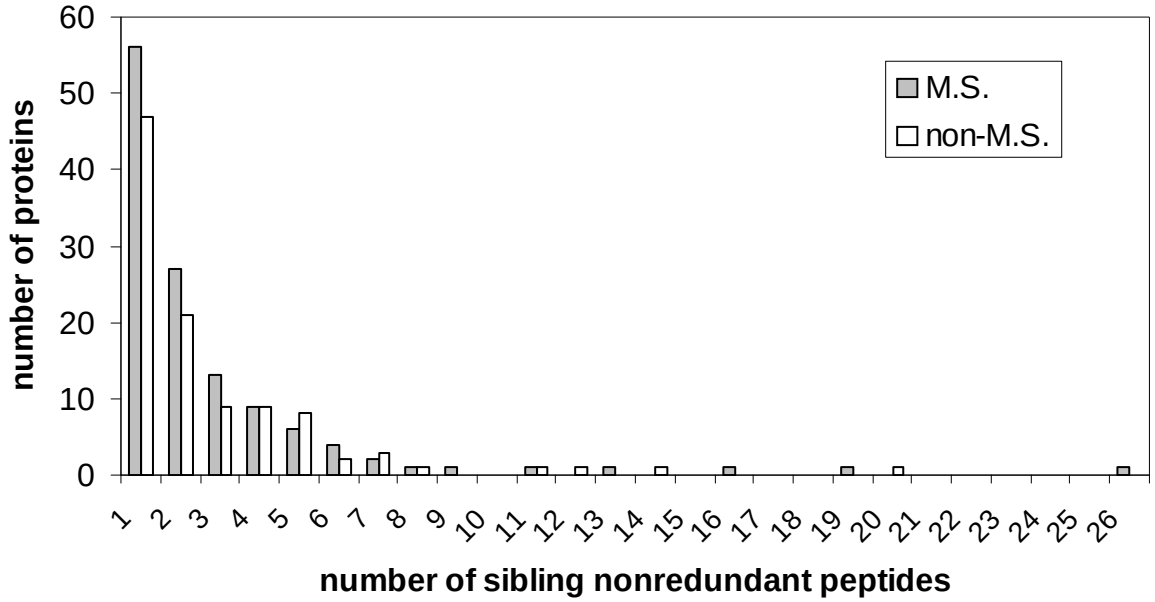

Figure 3: Number of proteins matched by a given number of peptides

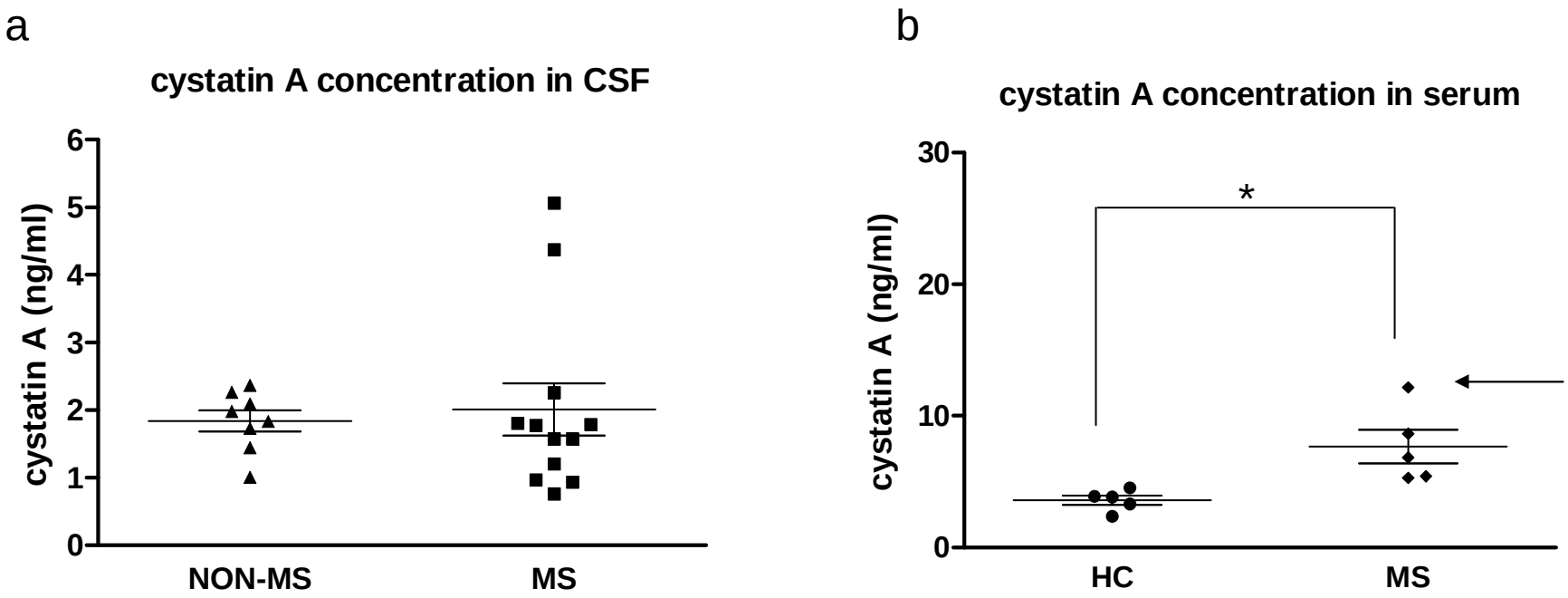

Figure 4: Cystatin A levels in CSF(a) and serum (b) of M.S. and non-M.S. patients

Table 1: Patient characteristics

A. M.S. group

Patient code	Sex	Age	Disease type[a]	Disease duration	EDSS score[b]	Relapse rate[c]	Treatment	Protein content (mg/ml)	IgG index
M.S.1	F	36	RR	3	2.5	3	None	0,48	Xxx
M.S.2	F	24	RR	2	1	1	None	0,25	Xxx
M.S.3	M	31	RR	3	3.5	2	None	0,04	Xxx
M.S.4	F	51	RR	1	3.5	2	None	0,19	Xxx
M.S.5	F	31	RR	6	1.5	2	None	0,27	Xxx
M.S.6	F	34	RR	3	1.5	1	None	0,11	Xxx
M.S.7	F	35	RR	3	2.5	2	None	0,32	Xxx
M.S.8	M	24	RR	1	1	1	None	0,37	Xxx

B. non-M.S. group

Patient code	Sex	Age	Diagnosis	Treatment	Justification analysis	CSF status[d]
Non-M.S.1	F	27	leukemia	chemotherapy	cerebral metastasis?	normal
Non-M.S.2	M	59	testis carcinoma	chemotherapy	cerebral metastasis?	normal
Non-M.S.3	F	23	mamma carcinoma	chemotherapy	cerebral metastasis?	normal
Non-M.S.4	M	46	colon carcinoma	chemotherapy	leptomeningeal metastasis?	normal
Non-M.S.5	F	79	mamma carcinoma	chemotherapy	leptomeningeal metastasis?	normal
Non-M.S.6	F	53	mamma carcinoma	chemotherapy	cerebral metastasis?	normal

[a] RR: relapsing-remitting
[b] EDSS: expanded disability status scale
[c] Relapse rate: number of relapses in a period of two years prior to CSF collection
[d] Examined CSF parameters include liquor colour, leukocyte cell count, bacterial growth and presence of malignant cells

Table 2: Evolution of peptide and protein count number following a stepwise evaluation procedure

	Validation step	M.S.		non-M.S.	
		#peptides	#proteins	#peptides	#proteins
1	Sequest	604	237	407	137
2	Mascot	262[a]	107	219[a]	83
3	Scaffold	350[b]	124[c]	285[b]	104[c]

[a] significance threshold $p \leq 0.05$
[b] peptide identification probability: 95%
[c] protein identification probability > 91%

Table 3: Proteins identified in common (a) or **shotgunned in** the CSF of non-M.S. (b) and M.S. (c) patients

Table 3 a (1/ 4) **Proteins in common in M.S. and non-M.S. group**

#	Uniprot Code	Locus	Description	MW[a]	M.S.[b]	non-M.S.[c]	∩[d]	References
1	P07108	ACBP	Acyl-CoA-binding protein	9913	3	2	2	56,57
2	P05067	A4	Amyloid beta A4 protein	86943	3	2	2	25,40
3	P02760	AMBP	AMBP protein	38999	1[f]	2	0	26,53,48
4	P51693	APLP1	Amyloid-like protein 1	72176	5	3	2	59,40
5	Q06481	APLP2	Amyloid-like protein 2	86956	1	1	1	40
6	P02649	APOE	Apolipoprotein E	36154	2[f]	2	2	15,40,53,54,60
7	P01884	B2MG	Beta-2-microglobulin	13715	4	3	3	11,15,24,25,40,53,60
8	P80723	BASP	Brain acid soluble protein 1	22562	5[f]	2	2	49
9	P55290	CAD13	Cadherin-13	78287	3	3	3	40
10	Q9BRK5	CAB45	Calcium-binding protein 45 kDa	41807	1	1	1	
11	P13987	CD59	CD59 glycoprotein	14177	4	3	3	40
12	Q99674	CGRE 1	Cell growth regulator with EF hand domain 1	31977	3	1	1	40
13	P10645	CMGA	Chromogranin A	50730	13	12	10	24,25,40,53
14	P10909	CLUS	Clusterin (Apolipoprotein J)	52495	2	4	2	11,15,40,53,61
15	Q16568	CART	Cocaine- + amphetamine-regulated transcript	12829	1[f]	1	1	
16	P02452	C01A1	Collagen alpha 1(I) chain	138883	2	2	2	53,54
17	P01028	CO4	Complement C4	192771	2	2	2	15,25,40,53,61
18	P23582	ANFC	C-type natriuretic peptide	13246	1	1	1	40
19	P01034	CYTC	Cystatin C (Gamma-trace)	15799	4	5	4	15,25,40,53,59,60
20	P81605	DCD	Dermcidin	11284	2	1	1	15,40

Table 3 a (2/4) **Proteins in common in M.S. and non-M.S. group**

#	Uniprot Code	Locus	Description	MW[a]	M.S.[b]	non-M.S.[c]	∩[d]	References
21	Q9UBP4	DKK3	Dickkopf related protein-3	38291	2	2	2	40
22	Q12805	FBLN3	Fibulin-3	54641	1	1	1	25,40,62
23	O60883	ETBR2	Endothelin B receptor-like protein-2	52829	2	2	2	40
24	Q15668	NPC2	Epididymal secretory protein E1	16570	4	4	3	15,25,40
25	Q92520	FAM3C	FAM3C	24680	2[e]	1	1	15,16,40
26	P02671	FIBA	Fibrinogen alpha/alpha-E chain	94973	3	14	3	15,23,40
27	P06396	GELS	Gelsolin, plasma (brevin)	85698	1[f]	1	0	15,25,40,53,60
28	Q8NBJ4	GP73	Golgi Phoshoprotein 2	45333	1	2	1	
29	-	-	Hypothetical protein MGC15730 gil14043256	52202	1[e]	1	1	
30	P01834	KAC	Ig kappa chain C region	11609	5	1	1	15,40,60
31	P18065	IBP2	Insulin-like growth factor binding protein 2	35138	2	4	2	40
32	P22692	IBP4	Insulin-like growth factor binding protein 4	27934	3[f]	1	1	40
33	P24593	IBP5	Insulin-like growth factor binding protein 5	30570	1[f]	1	1	
34	P24592	IBP6	Insulin-like growth factor binding protein 6	25322	3	5	3	24,25,40
35	Q16270	IBP7	Insulin-like growth factor binding protein 7	29130	5	4	4	40,54
36	Q92876	KLK6	Kallikrein 6	26856	1[e]	1	0	15,40,53,54,61
37	Q9BZG9	LYNX1	Ly-6/neurotoxin-like protein 1	12641	1	1	1	40
38	O94772	LY6H	Lymphocyte antigen Ly-6H	14669	2[f]	2	2	40
39	P04156	PRIO	Major prion protein	27661	4	5	4	25,40

Table 3 a (3/4) **Proteins in common in M.S. and non-M.S. group**

#	Uniprot Code	Locus	Description	MW[a]	M.S.[b]	non-M.S.[c]	∩[d]	References
40	P25713	MT3	Metallothionein-III	6927	2	2	2	
41	Q8TCZ1	Q8TCZ1	MIC2L1 isoformE3-E4	22833	1[e]	1	1	
42	P20774	MIME	Mimecan (Osteoglycin)	33922	2	1	1	40
43	Q16653	MOG	Myelin-oligodendrocyte glycoprotein	28179	1	1	1	40
44	P13591	NCA11	Neural cell adhesion molecule 1, 140 kDa (N-CAM140)	93361	5	5	4	25,40,53
45	Q9NQX5	NPDC1	Neural proliferation differentiation and control protein-1	34516	1	2	1	25,40
46	P19022	CADH2	Neural-cadherin	99851	4[f]	5	4	40,54
47	O14594	PGCN	Neurocan core protein	142973	2	1	1	40
48	P05408	7B2	Neuroendocrine protein 7B2	23730	2	2	2	24, 25,40
49	O95502	NPTXR	Neuronal pentraxin receptor	52718	1	1	1	25,40
50	O15240	VGF	Neurosecretory protein VGF	67287	7	5	4	24,40
51	P48745	NOV	NOV protein homolog	39162	4	4	4	63
52	P10451	OSTP	Osteopontin	35423	9	8	8	24,25,40,54
53	P20472	PRVA	min alpha	11928	4	3	3	64
54	P98160	PGBM	Perlecan	468825	1[f]	1	0	40,53
55	P30086	PEBP	Phosphatidylethanolamine-binding protein	20926	3	7	3	15,25,59
56	P07602	SAP	Proactivator polypeptide	58113	6	5	5	40
57	O15354	GPR37	Probable G protein-coupled receptor GPR37	67114	2[e]	2	2	
58	P01210	PENK	Proenkephalin A	30787	4	3	2	24,40

Table 3 a (4/4) **Proteins in common in M.S. and non-M.S. group**

#	Uniprot Code	Locus	Description	MW[a]	M.S.[b]	non-M.S.[c]	∩[d]	References
59	P41222	PTGDS	Prostaglandin-H2 D-isomerase	21029	6	7	4	11,15,24,25,40,43,50
60	P23468	PTPRD	Protein-tyrosine phosphatase delta (R-PTP-D)	214760	2	1	1	
61	Q16849	PTPRN	Protein-tyrosine phosphatase-like N (R-PTP-N)	105848	1[g]	2	1	
62	Q13332	PTPRS	Protein-tyrosine phosphatase S (R-PTP-S)	217081	2	4	2	
63	Q15293	RCN1	Reticulocalbin	38890	1[f]	1[h]	1	54
64	P07998	RNAS1	Ribonuclease pancreatic (RNase 1)	17644	3	4	3	24,40
65	P05060	SCG1	Secretogranin I (Chromogranin B)	78246	26	20	17	24,25,40,49
66	P13521	SCG2	Secretogranin II (Chromogranin C)	70869	6	1	1	40
67	Q8WXD2	SCG3	Secretogranin III	52977	6	7	5	25,40
68	P02768	ALBU	Serum albumin	69367	11	5	4	11,15,24,40,43
69	P09486	SPR1	SPARC (osteonectin)	34632	3[f]	1[h]	1	40
70	Q14515	SPRL1	SPARC-like protein (hevin)	75216	16	11	11	16,25,40
71	P00441	SODC	Superoxide dismutase [Cu-Zn]	15805	2	2	2	43,51
72	Q08629	TIC1	Testican-1 (SPOCK protein)	49124	1	1	1	
73	Q92563	TIC2	Testican-2 (SPOCK protein 2)	46779	1	1	0	24,40
74	P10599	THIO	Thioredoxin	11606	3	2	1	
75	P04216	THY1	Thy-1 membrane glycoprotein	17935	1	2	1	25,40
76	O43493	TGON2	Trans-Golgi network integral membrane protein 2	51007	5	4	4	
77	P02766	TTHY	Transthyretin	15887	7	6	6	15,24,25,40,44,50
78	P02248	UBIQ	Ubiquitin	8565	8	6	6	11,24,43
79	Q9UPU3	SORC3	VPS10 domain-containing receptor SorCS3	135755	3	3	3	
80	Q14508	WFDC2	WAP four-disulfide core domain protein 2	12993	4	3	3	
			Total peptide count		285	259	205	

Table 3 b **Proteins shotgunned in non-M.S. group**

#	Uniprot Code	Uniprot Locus	Description	MW[a]	non-M.S.[c]	References
1	P14621	ACYP2	Acylphosphatase, muscle type isozyme	11008	1[i]	
2	P02763	A1AG1	Alpha-1-acid glycoprotein 1 (Orosomucoid 1)	23512	1[i]	25,40,53,59,60,65,66
3	P48539	PEP19	Brain-specific polypeptide PEP-19	6660	1[i]	
4	P39059	COFA1	Collagen alpha 1(XV) chain precursor	141930	1[i]	53
5	P05937	CALB1	Calbindin (vitamin D-dependent calcium binding protein)	29894	1	67
6	P00915	CAH1	Carbonic anhydrase I	28739	1	54
7	Q02487	DSC2	Desmocollin 2A/2B	99962	1	
8	Q14118	DAG1	Dystroglycan	97581	1[i]	40
9	P24534	EF1B	Elongation factor 1-beta (EF-1 beta)	24633	1	
10	P22794	EVI2A	EVI2A protein	25768	1	
11	P62942	FKB1A	FK506-binding protein1A (Peptidyl-prolyl cis-trans isomerase)	11820	1[i]	68
12	Q08380	LG3BP	Galectin-3 binding protein	65331	1[i]	
13	O94910	LPHN1	Latrophilin 1(Lectomedin-2)	162717	3	40
14	P61626	LYSC	Lysozyme C	16537	1[i]	15,61
15	Q9H8J5	MANS1	Mansc domain containing protein 1	46810	1[i]	
16	P58546	MTPN	Myotrophin (V-1 protein)	12764	1[h]	59
17	P29966	MACS	Myristoylated alanine-rich C-kinase substrate (MARCKS)	31413	2	59
18	O95158	NXPH4	Neurexophilin 4 precursor	33065	1[h]	
19	Q9Y4L1	OXRP	Oxygen-regulated protein 150 kDa	111335	1	
20	P04085	PDGFA	Platelet-derived growth factor, A chain precursor	24043	1[h]	40
21	P00747	PLMN	Plasminogen	90569	2	25,40,53,54
22	Q8WWX9	SELM	Selenoprotein M	16185	1[i]	
23	O75368	SH3L1	SH3 domain-binding glutamic acid-rich-like protein	12774	3	
24	P61278	SMS	Somatostatin	12736	1[i]	40
			Total peptide count		30	

Table 3 c (1/2) **Proteins shotgunned in M.S. group**

#	Uniprot Code	Locus	Description	MW[a]	M.S.[b]	References
1	P29312	1433Z	14-3-3 protein zeta/delta (KCIP-1)	27745	1[f]	12,64 ,69-72
2	P01009	A1AT	Alpha-1-antitrypsin precursor	46737	1[f]	11,16,26,53,54,59,60,73
3	P04083	ANXA1	Annexin A1 (Annexin I) (Lipocortin I)	38583	1[f,i]	15
4	P07355	ANXA2	Annexin A2 (Annexin II) (Lipocortin II)	38473	2[f]	64,67
5	P08758	ANXA5	Annexin A5 (Annexin V) (Lipocortin V)	35806	1[f,l]	64,67,74
6	P02593	CALM	Calmodulin	16706	2[f]	75
7	Q9NZT1	CALL5	Calmodulin-like protein 5	15921	1[g,l]	
8	P31944	CASP1	Caspase-14	27680	2[g]	
9	P12111	CO6A3	Collagen alpha 3 (VI) chain	343552	1[f,l]	
10	P01024	CO3	Complement C3	187164	1[i]	15
11	P29279	CTGF	Connective tissue growth factor	38069	1[f]	40
12	Q12860	CNTN1	Contactin 1	113320	1[i]	53
13	P12277	KCRB	Creatine kinase, B chain	42644	2[f]	64,67,76-78
14	P01040	CYTA	Cystatin A (Stefin A)	11006	2	15
15	Q9NYX4	D1IP	D1 dopamine receptor-interacting protein calcyon	23434	1[f,l]	
16	Q02413	DSG1	Desmoglein 1	113716	1[g,h]	
17	P36957	ODO2	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	48640	1[g,h]	64
18	P30040	ERP29	Endoplasmic reticulum protein ERp29	28993	1[f,l]	
19	P14138	EDN3	Endothelin-3 precursor	25454	1[f,l]	79-83
20	P11021	GRP78	Glucose-regulated protein 78 kDa (GRP 78)	72333	2[f]	64,67
21	P07900	HS90A	Heat shock protein HSP 90-alpha (HSP 86)	84543	1[f,l]	64
22	P08238	HS90B	Heat shock protein HSP 90-beta (HSP 84) (HSP 90)	83133	1[f,l]	64
23	-	HGFL	Hepatocyte growth factor-like protein (HGFL) gi:15029677	28672	2	
24	P01617	KV2D	Ig kappa chain V-II region TEW	12316	1[e,h]	64

Table 3 c (2/2) **Proteins shotgunned in M.S. group**

#	Uniprot accession n°	Locus	Description	MW[a]	M.S.[b]	References
25	P01842	LAC	Ig lambda chain C regions	11237	1[e,l]	40,54
26	O75071	K0494	KIAA0494	55031	1[f,l]	
27	P33908	MA1A1	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	73068	1[f]	40
28	Q9P2S2	NRX2A	Neurexin 2-alpha precursor	184982	1[f]	
29	P41271	NBL1	Neuroblastoma suppressor of tumorigenicity 1	19277	1[e,i]	40
30	P17677	NEUM	Neuromodulin (Calmodulin-binding protein P-57)	24803	1[f,l]	84
31	Q02818	NUCB1	Nucleobindin 1	53821	1[f,i]	
32	P19021	AMD	Peptidyl-glycine alpha-amidating monooxygenase (PAM)	108332	1[e,h]	54
33	Q9UHG2	Q9UHG2	ProSAAS	27372	1[e,i]	40,53
34	P49221	TGM4	Protein-glutamine glutamyltransferase 4 (TGase 4)	77145	1	
35	P31151	S10A7	Psoriasin	11326	1[e]	15
36	P49792	RBP2	Ran-binding protein 2 (Ran BP2)	358218	1[g,i]	54
37	P23471	PTPRZ	Receptor-type protein-tyrosine phosphatase zeta (R-PTP-Z)	254530	1[f,i]	40
38	P35268	RL22	Ribosomal protein L22 60S	14656	1[f,h]	
39	P02787	TRFE	Serotransferrin (Transferrin)	77050	19	15
40	Q9H299	SH3L3	SH3 domain-binding glutamic acid-rich-like protein 3	10438	1[f]	
41	Q9UPR5	NAC2	Sodium/calcium exchanger 2	100368	1[f]	
42	P14209	MIC2	T-cell surface glycoprotein E2 (CD99 antigen)	18848	2	25,40
43	P55327	TPD52	Tumor protein D52	19863	1[f]	
44	P08670	VIME	Vimentin	53555	2[f]	85,86
			Total peptide count		71	

[a] theoretical molecular weight of the protein precursor form

[b] number of sibling peptides matching the identified protein in the M.S. group

[c] number of sibling peptides matching the identified protein in the non-M.S. group

[d] number of common sibling peptides matching the identified protein in the M.S. and non-M.S. group

[e] proteins identified in the M.S. group by GPF_M.S.A only

[f] proteins identified in SCX_1min_M.S.B only

[g] proteins identified in SCX_2min_M.S.A only

[h] single-peptide protein identifications evaluated by Lutefisk for which no quality *de novo* sequences could be derived or for which the *de novo* sequences differed from the one found by Sequest and Mascot

[l] single-peptide protein identifications shotgunned in the M.S. or non-M.S. group for which a p value < 0.01 was reached

Table 4: Proteins not previously identified by mass spectrometry in CSF

Acyl-CoA-binding protein [a]	KIAA0494
Acylphosphatase	Mansc domain containing
Annexin A2 [b]	Metallothionein-III
Annexin A5 [a,b]	MIC2L1 isoformE3-E4
Brain-specific polypeptide PEP-19	Myotrophin
Calbindin [b]	Neurexin-2 alpha
Calcium-binding protein 45 kDa	Neuromodulin [a]
Calmodulin [a]	Neurexophilin 4
Calmodulin-like protein 5	NOV protein homolog [a]
Caspase-14	Nucleobindin 1
Cocaine- and amphetamine-regulated transcript protein	Oxygen-regulated protein 150 kDa (Hypoxia up-regulated 1)
Collagen alpha 3	Parvalbumin [b]
Creatine kinase B chain [a,b]	Probable G protein-coupled receptor GPR37
D1 dopamine receptor-interacting protein calcyon	Protein glutamine glutarymyltransferase 4
Desmocollin 2A/2B	Protein-tyrosine phosphatase delta (R-PTP-D)
Desmoglein 1	Protein-tyrosine phosphatase-like N (R-PTP-N)
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial [b]	Receptor-type protein-tyrosine phosphatase S (R-PTP-S)
Elongation factor 1-beta	Reticulocalbin [b]
Endoplasmic reticulum protein ERp29	Ribosomal protein L22 (60S)
Endothelin 3 [a]	Selenoprotein M
EVI2A	SH3 domain-binding glutamic acid-rich-like protein
Galectin-3 binding protein	Sodium/calcium exchanger 2
Glucose-regulated protein 78 [b]	Testican-1
Golgi phosphoprotein 2	Thioredoxin
Hepatocyte growth factor-like protein	Trans-Golgi network integral membrane protein 2
Heat shock protein HSP 90-alpha (HSP 86) [b]	Tumor protein D52
Heat shock protein HSP 90-beta (HSP 84) (HSP 90) [b]	Vimentin [a,b]
MGC 15730	VPS10 domain-containing receptor SorCS3
Ig kappa chain V-II region TEW [b]	WAP four-disulfide core domain protein 2
Insulin-like growth factor binding protein 5	14-3-3 zeta [a,b]

[a] proteins reported in human CSF using non-mass spectrometry-based studies (for references see table 3)
[b] proteins reported in human brain proteomic studies (for references see table 3)

Figure 1: Number of tandem mass spectrometric files recorded at a given SCX fraction collection time

Pooled and ultrafiltered CSF samples of M.S. patients were digested with trypsin and separated on the SCX column. The M.S. pool was collected either every two minutes (SCX_2min_M.S.A; 26 fractions, 5 µl each) or every minute (SCX_1min_M.S.B; 55 fractions, 2.5 µl each). The SCX fractions of pool 'non-M.S.' comprising CSF of 6 cancer patients were collected every minute (SCX_1min_non-M.S.; 55 fractions, 2.5 µl each). The number of mass spectrometric files recorded for each fraction was set equal to the number of dta files generated by Sequest from the Xcalibur raw data files.

Figure 2: Number of protein identifications for each fractionation method used on CSF of M.S. patients

Pooled and ultrafiltered CSF samples of M.S. patients were digested with trypsin and separated on the SCX column. Pool M.S.B. was collected each minute (SCX_1min_M.S.B; 55 fractions, 2.5 µl each) while pool M.S.A was collected every two minutes (SCX_2min_M.S.A; 26 fractions, 5 µl each). As an alternative, the digested subproteome of M.S.A was analyzed directly by LC-ESI-MS/MS and gas-phase fractionation (GPF-M.S.A) in the mass spectrometer. Combining SCX_1min_M.S.B and GPF_M.S.A experiments, 118 out of the 124 proteins (95 %) were covered for the M.S. patient group.

Figure 3: Number of proteins matched by a given number of peptides

Pooled and ultrafiltered CSF samples of M.S. and non-M.S. patients were digested with trypsin and subjected to SCX and GPF. Mass spectrometric data were validated in a stepwise matter (Table 2) resulting in 148 confident protein identications. Single-peptide protein identifications amounted to 45% in both the M.S. and non-M.S. group.

Figure 4: Cystatin A levels in serum and CSF of M.S. and non-M.S. patients

A cystatin A-specific ELISA was performed on CSF (a) and serum samples (b) of M.S. and non-M.S. patients. (a) Hundred µl of CSF from 12 M.S. patients (all RR in state of remission) and 8 non-M.S. patients (2 breast tumor patients and 6 patients with non-inflammatory neurological disorders) was used to determine the cystatin A concentration. (b) Hundred µl of diluted serum samples of 6 M.S. patients (all RR in state of remission) and 6 healthy controls were used. *There was a significant ($p < 0.01$) increase of cystatin A level in serum of M.S. patients compared to healthy controls. Paired data points available for 1 M.S. patient are indicated with an arrow.

Synopsis

Lumbar cerebrospinal fluid proteome in multiple sclerosis: characterization by ultrafiltration, liquid chromatography and mass spectrometry

Jean-Paul Noben, Debora Dumont, Natalia Kwasnikowska, Peter Verhaert, Veerle Somers, Raymond Hupperts, Piet Stinissen and Johan Robben

Pooled and ultrafiltered CSF of multiple sclerosis (M.S.) and non-M.S. patients were analyzed by 2D-LC-ESI-MS/MS and 1D-LC-ESI-MS/MS, the latter with gas-phase fractionation (GPF) in the iontrap.These proteomic approaches in combination with a three-step evaluation process resulted in the identification of in total 148 proteins, including 60 proteins evidenced here for the first time in CSF by mass spectrometry.