Expanding the cerebrospinal fluid endopeptidome

Karl T Hansson[1], Tobias Skillbäck[1,2], Elin Pernevik[1], Silke Kern[1,2], Erik Portelius[1,2], Kina Höglund[1,2], Gunnar Brinkmalm[1], Jessica Holmén-Larsson[1], Kaj Blennow[1,2], Henrik Zetterberg[1,2,3], Johan Gobom[1,2]

[1]Inst. of Neuroscience and Physiology, Dept. of Psychiatry and Neurochemistry, The Sahlgrenska Academy at University of Gothenburg, Mölndal, Sweden

[2]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

[3]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, UK

Corresponding author:

Johan Gobom

Clinical Neurochemistry Laboratory

Sahlgrenska University Hospital

University Hospital

SE-431 80 Mölndal

Sweden

Email: johan.gobom@neuro.gu.se

Abstract

Biomarkers of neurodegenerative disorders are needed to assist in diagnosis, to monitor disease progression and therapeutic interventions, and to provide insight into disease mechanisms. One route to identify such biomarkers is by proteomic and peptidomic analysis of cerebrospinal fluid (CSF).

In the current study, we performed an in-depth analysis of the human CSF endopeptidome to establish an inventory that may serve as a basis for future targeted biomarker studies. High-pH reversed-phase HPLC was employed for peptide fractionation followed by low-pH nano-LC-MS analysis. Different software programs and scoring algorithms for peptide identification were employed and compared.

A total of 18,031 endogenous peptides were identified (FDR = 1%), increasing the number of known CSF peptides 10-fold compared to previous studies. The peptides were derived from 2,053 proteins of which more than 60 have been linked to neurodegeneration. Notably, among the findings were six peptides derived from microtubule-associated protein tau, three of which span the diagnostically interesting threonine-181. Also, 213 peptides from amyloid precursor protein (APP), 58 of which were partially or completely within the sequence of amyloid β 1-40/42, as well as 109 peptides from apolipoprotein E, spanning sequences that discriminate between the E2/E3/E4 isoforms of the protein.

Keywords

Cerebrospinal fluid, peptides, peptidomics, biomarkers, neurodegeneration, Alzheimer's disease, Parkinson's disease

Introduction

Analysis of cerebrospinal fluid (CSF) is valuable to study neurodegenerative disorders (1). Research on diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD) would benefit greatly from new biomarkers that can aid in diagnosis, be used for monitoring disease progression, and provide insight into the disease mechanisms. As new disease-modifying therapies are being developed, for example against AD (2), there will be an increased need for biomarkers that enable earlier and more accurate diagnosis, as well as to provide means to monitor disease progression and response to treatment.

Produced as an ultra-filtrate of blood in the ventricles and around the blood vessels of the central nervous system (CNS), CSF circulates around the brain and the spinal cord and is drained into the blood, with a turnover time of approximately 8 h (3). A multitude of molecules shed by cells in the brain are present in CSF, and thus, many processes in the CNS are dynamically reflected in the molecular composition of the CSF. Approximately 20% of the protein content is derived from the CNS through the interstitial fluid and the remainder can be traced to plasma constituents passing the blood-brain barrier (4-9). Shotgun proteomic studies of CSF have led to the identification of biomarker candidates of several neurodegenerative disorders, including AD, PD, and multiple sclerosis (10).

Studies by our group and others have revealed that CSF, besides proteins, contains many endogenous peptides (11-16). Their concentration being dependent on a variety of processes, such as enzymatic protein processing, secretion, and aggregation, these peptides may convey valuable biomarker information. Such information may obviously be lost when following the currently standard analytical strategy in proteomics of digesting the sample proteins with trypsin and basing protein quantification on measurement of tryptic peptides.

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

In a pilot study we identified quantitative differences in endogenous peptides in AD patients compared to controls (17), suggesting their potential as biomarkers. In a recent study, we used endopeptidomics to detect changes in the abundance of endogenous CSF peptides in healthy individuals following treatment with the γ-secreatase inhibitor semagacestat demonstrating target engagement (18).

In order to further investigate the potential of endogenous peptides in CSF as biomarkers, the aim of the current study was to expand the known CSF peptidome to include also more low-abundant peptides. Our approach was to implement a step of peptide pre-fractionation to allow increased CSF volume to be used for LC-MS analysis, and employing software that employs different algorithms for peptide identification and scoring of spectrum matches.

Experimental Procedures

Participants and sampling of cerebrospinal fluid

Cerebrospinal fluid was sampled from one patient diagnosed with hydrocephalus undergoing evaluation at Lund University Hospital, Lund, Sweden; as well as from CSF pooled from five individuals undergoing evaluation at the Clinical Neurochemistry Laboratory, Sahlgrenska

University Hospital, Mölndal, Sweden. Sampling was performed by lumbar puncture in accordance with a standardised protocol previously developed on site. Non-soluble material, cells and cell debris in the CSF were removed by centrifugation at 2000 x g and +4 °C for 10 min. Surplus CSF from patients undergoing clinical evaluation was used for this project after de-identification, as approved by the regional ethics committee at the University of Gothenburg.

Materials

Ammonium bicarbonate (AmBic) [40867-F], 25% ammonium hydroxide [30501], 1 M triethylammonium bicarbonate pH 8.5 (TEAB) [T7408], 8 M guanidinuim hydrochloride (Gua-HCl) [G9824], ≥98% sodium deoxycholic acid (Na-DOC) [30970], ≥98% *tris*(2-carboxyethyl)phosphine (TCEP) [C4706], ≥ 99% iodacetamide (IAA) [I1149], 99% trifluoracetic acid (TFA) [T6508] and 98% formic acid (FA) [56302-F] were all acquired from Sigma-Aldrich; HPLC-gradient grade Far-UV acetonitrile (AcN) [A998] and iso-propanol [A520] from Fisher Scientific; trypsin acquired from Promega [V5111]]; Amicon Ultra-15 Centrifugal Filter Units 30 kDa Molecular weight cut-off (MWCO) ultracentrifuge-filters [UFC903024] were obtained from Merck Millipore and Sep-Pak® Vac 1cc (100mg) C_{18} cartridges [WAT023590]was obtained from Waters.

CSF sample preparation – endogenous peptides

CSF peptide extracts were prepared according to a previously described protocol (19), modified to accommodate a larger sample volume. Briefly, 1.5 ml aliquots of CSF were thawed at room temperature (RT), vortexed gently, and transferred to 15 ml Falcon tubes. Aliquots of 1 M TEAB (250 µl) and 8 M GuaHCl (750 µl) were added and the samples were vortexed gently. Reduction of protein/peptide disulphides was attained by addition of a solution of 200 mM TCEP (60 µl), 300 mM TEAB (60 µl) followed by 1 h incubation at +55

°C under gentle agitation. The samples were subsequently alkylated by addition of 60 µl 400 mM IAA (aq) followed by 30 min incubation at RT in darkness. The samples were diluted by addition of 2.38 ml water (MilliQ) to a total volume of 5 ml and vortexed.

Isolation of endogenous peptides was performed by ultrafiltration using 30 kDa MWCO filter devices. The filters were conditioned by adding a solution of 100 mM TEAB, 3 M Gua-HCl (5 ml) and centrifuging at 2,500 x g for 20 min at RT, discarding the flow-through. The samples were then loaded and centrifuged (2,500 x g for 45 min, RT). To improve recovery, 50 mM AmBic (3 ml) were loaded, spun through (RT, 2,500 x g for 20 min) and pooled with the filtrate.

The sample was acidified (final pH ≈ 3) by titration with 0.1% TFA and subsequently desalted by SPE (SEP-Pak C_{18}), operated using a vacuum chamber. The cartridge was conditioned with 2 x 1 ml 84% AcN, 0.1% TFA, and equilibrated by 2 x 1 ml 0.1% TFA, after which the sample was loaded, washed with 2 x 1 ml 0.1% TFA and subsequently eluted with 1 ml 84% AcN, 0.1% TFA. Finally the eluate solvents were evaporated in a vacuum concentrator and at –80 °C.

CSF sample preparation – tryptic peptides

For comparing the identification rates of endogenous and tryptic peptides, an aliquot of the CSF sample from the hydrocephalus patient was digested with trypsin. CSF (100 µl) was thawed at RT, vortexed gently and transferred to a 1.5 ml Eppendorf tube. Aliquots of 1 M TEAB (16.7 µl) and 3.3% Na-DOC (50 µl) were added and the sample was vortexed. Reduction of protein disulphides was performed by adding an aliquot of 200 mM TCEP, 300 mM TEAB (4 µl) followed by 1 h incubation at +55 °C under gentle agitation. Cysteines were alkylated by addition of 400 mM IAA (4 µl) followed by 30 min incubation at RT in darkness. Trypsin (Promega, 20 µg) was dissolved in 50 mM acetic acid (100 µl) and pre-incubated for 15 min at 37 °C. A 25 µl aliquot of the solution was transferred to the sample,

which was incubated 12 h at 37 °C. The sample was subsequently desalted, concentrated, and stored as described above.

Alkaline Reverse-Phase HPLC Separation and Fraction Concatenation

HpH-RPLC fractionation was performed according to the method of Bath et al (20) with minor method alterations, using a Ultimate 3000 HPLC system (Thermo) equipped with an integrated fraction collector adapted for 96-well deep well plates. Endogenous peptide extracts (from 1.5 ml CSF) were dissolved in 2.5 mM NH_3(aq) and 2% AcN (15 µl), and separated over an XBridge Peptide BEH C_{18} Column, 130 Å . The mobile phases consisted of (A) pure water; (B) 84% AcN; and (C) 25 mM NH_3(aq). Buffer C was made fresh at the start of each trial and kept on ice to minimize NH_3(aq) decomposition/evaporation. The LC was operated at a constant flow of 100 µl/min with a constant concentration of 10% Buffer C, and using the following gradient: t (min)=0, %B=2%; t=4, B=2%; t=50, B=90%; t=66, B=90%; t=67, B=2%; t=76, B=2%. Fraction collection was started at t=4 min, collecting 72 one-minute fractions in a cycling pattern over 24 wells in a 96-well deep-well plate, resulting in 3 concatenated fractions per well. The fractions were further concatenated by combining the contents of well 1 and 13, 2 and 14, 3 and 15 etc., resulting in 12 samples, each containing six of the original fractions. The fractions were subsequently split in two aliquots, vacuum concentrated and stored at –80 °C until LC-MS analysis.

LC-MS

Chromatography was performed on an Ultimate 3000 RSLC nano system (Thermo) in trap column configuration (trap column: Acclaim® PepMap 100 (Thermo), 75 µm x 2 cm, C_{18}, 100 Å pore size, 3 µm particle size; separation column: Acclaim® PepMap C18, 75 µm x 500 mm, 100Å pore size, 2 µm particle size). The mobile phases were (A) 0.1% FA and (B) 84% AcN, 0.1% FA. Sample aliquots corresponding to 750 µl CSF were dissolved in 0.05% TFA,

2% AcN (6 µl) (loading buffer). 5 µl were loaded at 5 µl/min using 0.05% TFA, 2% AcN. The following gradient was used: t (min) =0, B=2%; t=10, B=2%; t=11, B=7%; t=100, B=26%; t=170, B=45%; t=175, B=80%; t=181, B=2%; t=210, B=2%. The LC was connected to an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo) via a FlexiSpray nano-ESI interface (Thermo). Full scan spectra were recorded in MS mode at a resolution setting of 120,000 (2.0e5 AGC target) over the *m/z* range 350-1400. The mass spectrometer was operated in the data-dependent acquisition mode, recording MS/MS spectra from the top ten most intense peaks with *m/z* > 150 and within the intensity range 1.0e4-1.0e5 were selected for fragment ion analysis. Precursor ions were isolated using a quadrupole isolation window of 3 *m/z*. Dynamic exclusion was used, with an exclusion time of 15 s and an *m/z* tolerance of ±10 ppm. Fragmentation was performed in the higher-collision energy dissociation (HCD) cell (29% collision energy) and MS/MS acquisitions were recorded in the orbitrap at a resolution setting of 30,000 (5.0e4 AGC target value). A replicate analysis was performed of each fraction.

Peptide identification

Peptide identification was performed using Mascot v2.4 (Matrix Science), SequestHT (Thermo), and PEAKS Studio v7.5 (Bioinformatic Solutions Inc.). The following settings were used for all three programs unless otherwise specified: database: UniProt_SwissProt [version2015_11]; taxonomy: homo sapiens; enzyme: none (trypsin for digested samples); max. missed cleavages: 0 (endogenous peptides), 2 (tryptic peptides); instrument (Mascot only): ESI-Trap; min. peptide length (SequestHT only): 6; precursor mass tolerance: 15 ppm; fragment mass tolerance: 0.05 Da; static modifications: Carbamidomethyl (C); dynamic modifications: Oxidation (M); peptide-spectrum match (PSM) validator: Target Decoy or Percolator (Mascot and Sequest HT only) or Decoy Fusion (PEAKS only); target FDR: 0.01.

Evaluation of the respective capacity of Mascot and Sequest HT to identify peptides from mass spectra was performed by comparing the results from two different scoring functions for PSM-validation; the Percolator algorithm (21, 22) and a regular target decoy (TD) validator (23), both implemented in the software Proteome Discoverer 2.0 (Thermo).

Confirmation of identification of tau endogenous peptides in CSF

Two synthetic, stable isotope-labelled peptides, NH_2-TPPAPKTPSSGEPP[K(^{13}C6)]-COOH and NH_2-TPPAPK[T(PO^{3}H2)]PSSGEPP[K(^{13}C6)]-COOH, corresponding to amino acids 175-190 of microtubule-associated protein tau, were custom synthesized (AQUA peptides; Thermo Scientific) to confirm the identification of their native counterparts. 50 fmol each of $_{175}NH_2$-TPPAPKTPSSGEPP[K(^{13}C6)]-$COOH_{190}$ and $_{175}NH_2$-TPPAPK[T(PO^{3}H2)]PSSGEPP[K(^{13}C6)]-$COOH_{190}$ were spiked into the endogenous sample corresponding to 1.5 ml CSF following pre-treatment but prior to fractionation.

Results and Discussion

To increase the number of identified endogenous peptides in CSF we supplemented a previously developed protocol (19, 24), based on molecular weight ultrafiltration, with an additional peptide fractionation step employing high-pH reversed-phase chromatography according to the method described by Batth and colleagues (20). Besides reducing sample complexity, pre-fractionation enabled us to use a fivefold larger volume of CSF for nano-LC-MS analysis compared to our previous studies, permitting the detection of lower-abundant peptides. We also evaluated software programs that use different strategies for peptide identification and validation of PSMs. Identified peptides were evaluated based on previously reported association to neurodegenerative disorders and peptides of particular interest (i.e., derived known or potential biomarkers) were examined further.

Peptide identification

The identification success rate is generally lower for endogenous peptides than for tryptic peptides. An LC-MS data set of tryptic peptides from CSF, consisting of 427,613 MS/MS

spectra, resulted in 36,886 confident (FDR<1%) PSMs, corresponding to 8.6% identified spectra, while a set of endogenous CSF peptides, consisting of 269,945 MS/MS spectra, yielded 3,094 accepted PSMs, i.e., only 1.1% identified spectra. A reason for the lower identification rate for endogenous peptides may be that they differ in composition compared to tryptic peptides: for example, they are on average longer, and contain lysine and arginine residues internally but not necessarily at the C-terminus. Thus PSM scoring algorithms optimized for tryptic peptides may not be optimal for endogenous peptides. To test this hypothesis, we searched LC-MS data from one sample (12 HpH-RPLC fractions) with the software programs Mascot (25, 26) and SequestHT (27, 28), which are based on fragment ion fingerprinting (26, 29), for both programs using either the default scoring algorithm or the Percolator algorithm (21, 30). While both algorithms use the target/decoy approach to assess the correctness of PSMs, Percolator employs machine learning to improve the base scoring algorithm based on a subset of highly confident PSMs in an iterative process. Thereby, the algorithm can adapt to the general fragmentation characteristics of peptides in the given sample and mass spectrometric experiment. Mascot identified 1,276 endogenous peptides using the default algorithm, and 8,679 peptides using Percolator, corresponding to over a 6-fold increase (Figure 1A). Sequest HT identified 1,694 endogenous peptides using its default scoring algorithm, and improved more than four-fold, to 7,288 peptides when instead applying Percolator (Figure 1B). For tryptic peptides the differences between Percolator and the default scoring algorithms were smaller: Mascot identified 5,824 peptides using the default algorithm and 7,926 peptides using Percolator, corresponding to an increase of 36% (Figure 1C), and SequestHT 7,028 peptides using the default algorithm and 7,917 using Percolator, corresponding to an increase of 12% (Figure 1D).

Another difference regarding identification of endogenous peptides compared to tryptic peptides is that no enzymatic cleavage can be specified in the database search. Thereby, the

number of peptide sequences to evaluate increases by a factor of 100-1,000, increasing the occurrence of incorrect PSMs. *De novo* sequencing is an alternative to fragment ion fingerprinting for peptide identification, in which partial peptide sequences are extracted by using intrinsic *m/z* information in MS/MS spectra (31). Because this approach does not rely on knowledge of the amino acids at the peptide termini, we hypothesized that it may be more successful for identification of endogenous peptides. To test this hypothesis, we searched the LC-MS data from the previous section using PEAKS Studio (32), which identifies peptides by *de novo* sequencing followed by a tag search (33, 34), and uses a PSM validator which combines machine learning and target decoy functions, some of which are also featured in Percolator, in a *decoy fusion* PSM validator (35). PEAKS identified 10,967 peptides (1% FDR), i.e., nearly 2,500 peptides more than Mascot and more than 3,000 peptides more than SequestHT (Figure 2).

The identification overlap between PEAKS and the data-dependent search engines was quite small. Compared to 70% peptide identification conformity between Mascot and Sequest HT, the identification overlap between PEAKS and either of the other two search engines was less than 20%, and the total overlap between all three was less than 15% (Figure 2). The number of peptides identified with the different programs and scoring algorithms are listed in Table 1.

The small overlap between the fingerprinting and *de novo* based methods raises the question of whether the identifications are correct or not. While correctness cannot be directly tested in this data set, for the subset of MS/MS spectra with assigned PSMs in at least two programs, the percent discrepant PSM assignments between the programs could reveal if the actual FDR of either of the programs is higher than the target value. The discrepancies between the programs were below 2% in all cases, thus not indicating erroneous FDR (Table 2).

In conclusion, for identification of endogenous peptides by fragment ion fingerprinting, the optimal scoring algorithm differs from that of tryptic peptides, and a significant improvement

can be achieved by using an adaptive scoring approach such as Percolator. Further, our results show that peptide identification based on *de novo* sequencing and fragment ion fingerprinting are complementary and should be used in combination for increased number of identified peptides.

Evaluation of biomarker candidates

Combining the results from the two CSF samples searched by three search engines resulted in a total of 18,031 unique endogenous peptides (FDR=1%), derived from 1,918 proteins. The set of proteins identified in our study was compared to CSF biomarkers and biomarker candidates presented in four recent review articles (3, 10, 36, 37) (see Table 3).

Microtubule-Associated Protein Tau

Among the identified endogenous peptides of particular interest were five derived from microtubule-associated protein tau (Table 4, Figure 3). In humans the primary role of tau is to stabilise neuronal microtubules, which are vital to axonal transport (38, 39). Intraneuronal aggregation of tau in the brain is involved in several neurodegenerative diseases, which are collectively referred to as *tauopathies* (40, 41). A specific pathophysiological trait of AD is phosphorylation of one or several tau-motifs (42-44), resulting in so-called hyperphosphorylated tau (p-tau). Hyperphosphorylation causes tau to dissociate from the microtubule-network, leading to its breakdown, and eventually results in cell-death (44, 45), resulting in increased CSF tau level (46-48). Together with the amyloid beta 1-42 peptide ($A\beta_{1-42}$), total tau (t-tau) and p-tau can be used to diagnose AD with high sensitivity and specificity (1, 37, 48, 49).

A study in which fractionated CSF proteins were analysed by Western blot demonstrated the existence of a multitude of tau protein fragments spanning a wide molecular weight range (47). The data indicated differences in the abundance of several fragments between AD

patients and controls, suggesting that it may be important to study the disease association of specific tau fragments or endogenous tau peptides. Our study is the first to report endogenous tau peptides in human CSF. Interestingly, four of the identified peptides (tau 175-189, tau 175-190, tau 175-193 and tau 176-190) span Thr-181, which is the phosphorylation site detected by the most commonly used immunoassay for p-tau (50). The identity of tau 175-190 was verified by spiking a CSF sample with custom synthesized tau 175-190 and analysing by LC-MS (Figure 4). Additionally, tau 175-190 with phosphorylated Thr-181 was spiked in, demonstrating also the presence of the phosphorylated form in CSF (data not shown). Further studies are under way to evaluate the performance of tau 175-190 and p-tau 175-190 as biomarkers of AD and other tauopathies.

Amyloid β

A total of 213 peptides belonging to the amyloid precursor protein (APP) were identified. Of these, 58 were located fully or partly within the $APP_{672-713}$ (APP770; identifier: P05067-1) sequence, corresponding to $A\beta_{1-42}$, the main constituent of amyloid plaques, one of the hallmark pathological lesions of AD (Supplemental Table s1). A majority of the truncated, endogenous Aβ peptide forms we report on here have been previously identified using a combination of immunoprecipitation (IP) and MS (51, 52). However; we identified a number of truncated segments not previously reported, in some cases likely because their sequences were outside the epitopes of the antibodies (12EF325, 6E10, and 4G8) used. A further three peptides (-10 to 9, -15 to 9 and -8 to 2) were not completely located within the Aβ sequence.

Endogenous peptides located within the $A\beta_{1-42}$ sequence are of particular interest when studying enzymatic processing of the Aβ/APP, e.g., to confirm target engagement of beta-secretase inhibitor treatment (53)

Apolipoprotein E

109 endogenous peptides derived from apoE were identified (Supplemental Table s1). Apolipoprotein E (apoE) is a glycoprotein which known functions involve CNS lipid transport and peripheral circulation (54). It is also believed to be a "progenitor" for bioactive polypeptides resulting from proteolytic cleavage of the protein (55-57). ApoE is known to be involved in AD and other neurodegenerative diseases (58-60), and there is evidence suggesting the protein affects AD through interaction with $A\hat{a}_{1-42}$ (61, 62), but the exact mechanism is still not fully understood.

There are three major isoforms of apoE in humans which differ by two single-amino acid substitutions: å2 - Cys_{130}/Cys_{176}; å3 - Cys_{130}/Arg_{176} and å4 - Arg_{130}/Arg_{176}**.** Carrying at least one allele å4 of the *APOE* gene (*APOE*) correlates with an increased risk of developing AD (58, 60, 63), whilst å2 (APOE2) seems to have a neuroprotective function (62, 64), and å3 (APOE3) is neutral in terms of AD risk (60).

We identified 109 endogenous apoE peptides covering >70% of the protein-sequence. Among the identified peptides were 6 spanning AA-130, and 3 spanning AA-176 allowing for separate quantification of endogenous peptides produced from the different genetic variants in future studies.

Conclusions

The described sample preparation, peptide fractionation, and peptide identification using different search engines significantly expanded the known CSF peptidome. The large number of endogenous CSF peptides identified, many of which have been reported to be associated with neurodegenerative diseases and comprising also peptides from low-abundant proteins, strongly suggests that this class of molecules is a promising source of biomarkers to explore in clinical studies.

Acknowledgements

This work was supported by funding from the Swedish Research Council, Wallströms and Sjöbloms Stiftelse, Gun and Bertil Stohnes Stiftelse, Magnus Bergwalls Stiftelse, Åhléns-stiftelsen, Alzheimerfonden, Demensförbundet, Stiftelsen för Gamla Tjänarinnor, the Knut and Alice Wallenberg Foundation, Frimurarestiftelsen, and FoU-Västra Götalandsregionen.

Supporting Information

The following files are available free of charge:

Supplemental Table S1. All endogenous CSF peptides identified in this study.

References

1. Blennow, K.; Hampel, H.; Weiner, M.; Zetterberg, H., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol* 2010, 6, (3), 131-44.

2. Scheltens, P.; Blennow, K.; Breteler, M. M.; de Strooper, B.; Frisoni, G. B.; Salloway, S.; Van der Flier, W. M., Alzheimer's disease. *Lancet* 2016.

3. Hühmer, A. F.; Biringer, R. G.; Amato, H.; Fonteh, A. N.; Harrington, M. G., Protein Analysis in Human Cerebrospinal Fluid: Physiological Aspects, Current Progress and Future Challenges. *Disease Markers* 2006, 22, (1-2), 3-26.

4. Blennow, K.; Zetterberg, H., Understanding Biomarkers of Neurodegeneration: Ultrasensitive detection techniques pave the way for mechanistic understanding. *Nat Med* 2015, 21, (3), 217-219.

5. Filiou, M. D.; Turck, C. W.; Martins-de-Souza, D., Quantitative proteomics for investigating psychiatric disorders. *Proteomics Clin Appl* 2011, 5, (1-2), 38-49.

6. Guest, P. C.; Martins-de-Souza, D.; Schwarz, E.; Rahmoune, H.; Alsaif, M.; Tomasik, J.; Turck, C. W.; Bahn, S., Proteomic profiling in schizophrenia: enabling stratification for more effective treatment. *Genome Med* 2013, 5, (3), 25.

7. Guldbrandsen, A.; Vethe, H.; Farag, Y.; Oveland, E.; Garberg, H.; Berle, M.; Myhr, K.-M.; Opsahl, J. A.; Barsnes, H.; Berven, F. S., In-depth Characterization of the Cerebrospinal Fluid (CSF) Proteome Displayed Through the CSF Proteome Resource (CSF-PR). *Molecular & Cellular Proteomics* 2014, 13, (11), 3152-3163.

8. Kim-Cohen, J.; Caspi, A.; Moffitt, T. E.; Harrington, H.; Milne, B. J.; Poulton, R., Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry* 2003, 60, (7), 709-717.

9. Parker, C. E.; Pearson, T. W.; Anderson, N. L.; Borchers, C. H., Mass-spectrometry-based clinical proteomics – a review and prospective. *The Analyst* 2010, 135, (8), 1830-1838.

10. Kroksveen, A. C.; Opsahl, J. A.; Aye, T. T.; Ulvik, R. J.; Berven, F. S., Proteomics of human cerebrospinal fluid: Discovery and verification of biomarker candidates in neurodegenerative diseases using quantitative proteomics. *Journal of proteomics* 2011, 74, (4), 371-388.

11. Holtta, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome Analysis of Cerebrospinal Fluid by LC-MALDI MS. *PLoS One* 2012, 7, (8), e42555.

12. Jahn, H.; Wittke, S.; Zurbig, P.; Raedler, T. J.; Arlt, S.; Kellmann, M.; Mullen, W.; Eichenlaub, M.; Mischak, H.; Wiedemann, K., Peptide fingerprinting of Alzheimer's disease in cerebrospinal fluid: identification and prospective evaluation of new synaptic biomarkers. *PLoS One* 2011, 6, (10), e26540.

13. Schulz-Knappe, P.; Zucht, H. D.; Heine, G.; Jurgens, M.; Hess, R.; Schrader, M., Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb Chem High Throughput Screen* 2001, 4, (2), 207-17.

14. Wijte, D.; McDonnell, L. A.; Balog, C. I.; Bossers, K.; Deelder, A. M.; Swaab, D. F.; Verhaagen, J.; Mayboroda, O. A., A novel peptidomics approach to detect markers of Alzheimer's disease in cerebrospinal fluid. *Methods* 2012, 56, (4), 500-7.

15. Yuan, X.; Desiderio, D. M., Human cerebrospinal fluid peptidomics. *J Mass Spectrom* 2005, 40, (2), 176-81.

16. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J Proteome Res* 2008, 7, (1), 386-99.

17. Holtta, M.; Minthon, L.; Hansson, O.; Holmen-Larsson, J.; Pike, I.; Ward, M.; Kuhn, K.; Ruetschi, U.; Zetterberg, H.; Blennow, K.; Gobom, J., An integrated workflow for multiplex CSF proteomics and peptidomics-identification of candidate cerebrospinal fluid biomarkers of Alzheimer's disease. *J Proteome Res* 2015, 14, (2), 654-63.

18. Hölttä, M.; Dean, R. A.; Siemers, E.; Mawuenyega, K. G.; Sigurdson, W.; May, P. C.; Holtzman, D. M.; Portelius, E.; Zetterberg, H.; Bateman, R. J., A single dose of the γ-secretase inhibitor semagacestat alters the cerebrospinal fluid peptidome in humans. *Alzheimer's research & therapy* 2016, 8, (1), 1.

19. Hölttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* 2012, 7, (8), e42555.

20. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *Journal of Proteome Research* 2014, 13, (12), 6176-6186.

21. Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J., Accurate and Sensitive Peptide Identification with Mascot Percolator. *Journal of Proteome Research* 2009, 8, (6), 3176-3181.

22. Spivak, M.; Weston, J.; Bottou, L.; Käll, L.; Noble, W. S., Improvements to the Percolator Algorithm for Peptide Identification from Shotgun Proteomics Data Sets. *Journal of Proteome Research* 2009, 8, (7), 3737-3745.

23. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods* 2007, 4, (3), 207-214.

24. Hölttä, M., Quantitative neuroproteomics for biomarker discovery in Alzheimer's disease. 2014.

25. Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J., Comparison of Mascot and X!Tandem Performance for Low and High Accuracy Mass Spectrometry and the Development of an Adjusted Mascot Threshold. *Molecular & Cellular Proteomics* 2008, 7, (5), 962-970.

26. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20, (18), 3551-67.

27. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of The American Society for Mass Spectrometry* 1994, 5, (11), 976-989.

28. Dongré, A. R.; Eng, J. K.; Yates Iii, J. R., Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins. *Trends in Biotechnology* 1997, 15, (10), 418-425.

29. Henzel, W. J.; Watanabe, C.; Stults, J. T., Protein identification: the origins of peptide mass fingerprinting. *Journal of The American Society for Mass Spectrometry* 2003, 14, (9), 931-942.

30. Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S., Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *Journal of Proteome Research* 2008, 7, (1), 29-34.

31. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M., Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid communications in mass spectrometry* 1997, 11, (9), 1015-1024.

32. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G., PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* 2003, 17, (20), 2337-2342.

33. Mann, M.; Wilm, M., Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry* 1994, 66, (24), 4390-4399.

34. Mørtz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M., Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases. *Proceedings of the National Academy of Sciences* 1996, 93, (16), 8264-8267.

35. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B., PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics* 2011.

36. Blennow, K., Cerebrospinal Fluid Protein Biomarkers for Alzheimer's Disease. *NeuroRX* 2004, 1, (2), 213-225.

37. Höglund, K.; Fourier, A.; Perret-Liaudet, A.; Zetterberg, H.; Blennow, K.; Portelius, E., Alzheimer's disease—Recent biomarker developments in relation to updated diagnostic criteria. *Clinica Chimica Acta* 2015, 449, 3-8.

38. Drechsel, D. N.; Hyman, A.; Cobb, M. H.; Kirschner, M., Modulation of the dynamic instability of tubulin assembly by the microtubule-associated protein tau. *Molecular biology of the cell* 1992, 3, (10), 1141-1154.

39. Hirokawa, N.; Funakoshi, T.; Sato-Harada, R.; Kanai, Y., Selective stabilization of tau in axons and microtubule-associated protein 2C in cell bodies and dendrites contributes to polarized localization of cytoskeletal proteins in mature neurons. *The Journal of cell biology* 1996, 132, (4), 667-679.

40. Hernandez, F.; Avila, J., Tauopathies. *Cellular and Molecular Life Sciences* 2007, 64, (17), 2219-2233.

41. Mietelska-Porowska, A.; Wasik, U.; Goras, M.; Filipek, A.; Niewiadomska, G., Tau Protein Modifications and Interactions: Their Role in Function and Dysfunction. *International Journal of Molecular Sciences* 2014, 15, (3), 4671-4713.

42. Grundke-Iqbal, I.; Iqbal, K.; Tung, Y.-C.; Quinlan, M.; Wisniewski, H. M.; Binder, L. I., Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences* 1986, 83, (13), 4913-4917.

43. Hanger, D. P.; Anderton, B. H.; Noble, W., Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. *Trends in molecular medicine* 2009, 15, (3), 112-119.

44. Johnson, G. V.; Stoothoff, W. H., Tau phosphorylation in neuronal cell function and dysfunction. *Journal of cell science* 2004, 117, (24), 5721-5729.

45. Feinstein, S. C.; Wilson, L., Inability of tau to properly regulate neuronal microtubule dynamics: a loss-of-function mechanism by which tau might mediate neuronal cell death. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 2005, 1739, (2–3), 268-279.

46. Karch, C. M.; Jeng, A. T.; Goate, A. M., Extracellular Tau levels are influenced by variability in Tau that is associated with tauopathies. *Journal of Biological Chemistry* 2012, 287, (51), 42751-42762.

47. Meredith, J. E., Jr.; Sankaranarayanan, S.; Guss, V.; Lanzetti, A. J.; Berisha, F.; Neely, R. J.; Slemmon, J. R.; Portelius, E.; Zetterberg, H.; Blennow, K.; Soares, H.; Ahlijanian, M.; Albright, C. F., Characterization of novel CSF Tau and ptau biomarkers for Alzheimer's disease. *PLoS One* 2013, 8, (10), e76523.

48. Sjögren, M.; Vanderstichele, H.; Ågren, H.; Zachrisson, O.; Edsbagge, M.; Wikkelsø, C.; Skoog, I.; Wallin, A.; Wahlund, L.-O.; Marcusson, J.; Nägga, K.; Andreasen, N.; Davidsson, P.; Vanmechelen, E.; Blennow, K., Tau and Aβ42 in Cerebrospinal Fluid from Healthy Adults 21–93 Years of Age: Establishment of Reference Values. *Clin Chem* 2001, 47, (10), 1776-1781.

49. Fortea, J.; Vilaplana, E.; Alcolea, D.; Carmona-Iragui, M.; Sanchez-Saudinos, M. B.; Sala, I.; Anton-Aguirre, S.; Gonzalez, S.; Medrano, S.; Pegueroles, J.; Morenas, E.; Clarimon, J.; Blesa, R.; Lleo, A.; for the Alzheimer's Disease Neuroimaging, I., Cerebrospinal fluid beta-amyloid and phospho-tau biomarker interactions affecting brain structure in preclinical Alzheimer disease. *Ann Neurol* 2014.

50. Vanmechelen, E.; Vanderstichele, H.; Davidsson, P.; Van Kerschaver, E.; Van Der Perre, B.; Sjögren, M.; Andreasen, N.; Blennow, K., Quantification of tau phosphorylated at threonine 181 in human cerebrospinal fluid: a sandwich ELISA with a synthetic phosphopeptide for standardization. *Neuroscience Letters* 2000, 285, (1), 49-52.

51. Mustafiz, T.; Portelius, E.; Gustavsson, M. K.; Holtta, M.; Zetterberg, H.; Blennow, K.; Nordberg, A.; Lithner, C. U., Characterization of the brain beta-amyloid isoform pattern at different ages of Tg2576 mice. *Neurodegener Dis* 2011, 8, (5), 352-63.

52. Rogeberg, M.; Wettergreen, M.; Nilsson, L. N.; Fladby, T., Identification of amyloid beta mid-domain fragments in human cerebrospinal fluid. *Biochimie* 2015.

53. Portelius, E.; Dean, R. A.; Andreasson, U.; Mattsson, N.; Westerlund, A.; Olsson, M.; Demattos, R. B.; Racke, M. M.; Zetterberg, H.; May, P. C.; Blennow, K., beta-site amyloid precursor protein-cleaving enzyme 1(BACE1) inhibitor treatment induces Abeta5-X peptides through alternative amyloid precursor protein cleavage. *Alzheimers Res Ther* 2014, 6, (5-8), 75.

54. Ladu, M. J.; Reardon, C.; van Eldik, L.; Fagan, A. M.; Bu, G.; Holtzman, D.; Getz, G. S., Lipoproteins in the central nervous system. *Annals of the New York Academy of Sciences* 2000, 903, (1), 167-175.

55. Zhou, W.; Scott, S.; Shelton, S.; Crutcher, K., Cathepsin D-mediated proteolysis of apolipoprotein E: possible role in Alzheimer's disease. *Neuroscience* 2006, 143, (3), 689-701.

56. Elliott, D. A.; Tsoi, K.; Holinkova, S.; Chan, S. L.; Kim, W. S.; Halliday, G. M.; Rye, K.-A.; Garner, B., Isoform-specific proteolysis of apolipoprotein-E in the brain. *Neurobiol Aging* 2011, 32, (2), 257-271.

57. Wang, M.; Turko, I. V., Mass spectrometry quantification revealed accumulation of C-terminal fragment of apolipoprotein E in the Alzheimer's frontal cortex. *PLoS One* 2013, 8, (4), e61498.

58. Farrer, L. A.; Cupples, L. A.; Haines, J. L.; Hyman, B.; Kukull, W. A.; Mayeux, R.; Myers, R. H.; Pericak-Vance, M. A.; Risch, N.; van Duijn, C. M., Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *Jama* 1997, 278, (16), 1349-1356.

59. Martins-De-Souza, D.; Wobrock, T.; Zerr, I.; Schmitt, A.; Gawinecka, J.; Schneider-Axmann, T.; Falkai, P.; Turck, C. W., Different apolipoprotein E, apolipoprotein A1 and prostaglandin-H2 D-isomerase levels in cerebrospinal fluid of schizophrenia patients and healthy controls. *World J Biol Psychiatry* 2010, 11, (5), 719-28.

60. Kim, J.; Basak, J. M.; Holtzman, D. M., The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron* 2009, 63, (3), 287-303.

61. Tokuda, T.; Calero, M.; Matsubara, E.; Vidal, R.; Kumar, A.; Permanne, B.; Zlokovic, B.; Smith, J. D.; Ladu, M. J.; Rostagno, A., Lipidation of apolipoprotein E influences its isoform-specific interaction with Alzheimer's amyloid β peptides. *Biochemical Journal* 2000, 348, (2), 359-365.

62. Tiraboschi, P.; Hansen, L.; Masliah, E.; Alford, M.; Thal, L.; Corey-Bloom, J., Impact of APOE genotype on neuropathologic and neurochemical markers of Alzheimer disease. *Neurology* 2004, 62, (11), 1977-1983.

63. Bertram, L.; McQueen, M. B.; Mullin, K.; Blacker, D.; Tanzi, R. E., Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature genetics* 2007, 39, (1), 17-23.

64. Aono, M.; Bennett, E.; Kim, K.; Lynch, J.; Myers, J.; Pearlstein, R.; Warner, D.; Laskowitz, D., Protective effect of apolipoprotein E-mimetic peptides on N-methyl-D-aspartate excitotoxicity in primary rat neuronal–glial cell cultures. *Neuroscience* 2003, 116, (2), 437-445.

Tables

Table 1. Number of endogenous CSF peptides identified using Mascot and SequestHT with Percolator or the default scoring algorithms, or PEAKS.

Protein identification software	**Scoring algorithm**	**Number of identified peptides**
Mascot (endogenous / tryptic)	Default	1,276 / 5,824
	Percolator	8,679 / 7,926
Sequest HT (endogenous / tryptic)	Default	1,694 / 7,028
	Percolator	7,288 / 7,917
PEAKS (endogenous)	*De novo*-sequencing/ Target Decoy Fusion /	10,967

Table 2. Discrepancy between protein identification software.

Identification software compared	# Mutual peptide IDs	# Identical peptide ID/Scan	# Peptide ID discrepancies	Peptide ID discrepancy (%)
Mascot - PEAKS	18,337	17,984	353	1.93
SequestHT - PEAKS	16,108	16,044	64	0.40
Mascot - SequestHT	17,323	17,235	88	0.51

Table 3.

Protein	Acc. No.	Disease	# Peptides identified
Secretogranin-1	P05060	AD, PD (10)	681
Clusterin	P10909	AD, MS, PD (10)	318
Amyloid beta A4 protein	P05067	AD, CJD (3)	213
SeroTransferrin (Transferrin, Beta-1 metal-binding globulin)	P02787	AD, MS (3, 10)	189
SPARC-like protein 1	Q14515	AD, MS (10)	179
ProSAAS	Q9UHG2	AD, MS (10)	153
Cystatin-C	P01034	AD, PD (3, 10)	140
Prostaglandin-H2 D-isomeras (Beta-trace protein)	P41222	AD, MS, Spinal canal stenosis, Bacterial meningitis (3, 10)	123
Apolipoprotein E	P02649	AD, MS (3, 10)	109
Beta-2-microglobulin	P61769	AD, MS, PD (10)	96
Complement C3	P01024	AD, MS (10)	65
Neuronal pentraxin receptor	O95502	AD, MS, PD (10)	65
Complement C4-A	P0C0L4	AD, MS (10)	62
Superoxide dismutase [Cu-Zn]	P00441	AD, PD, MS (10)	61
Transthyretin (ATTR, Prealbumin)	P02766	AD, PD, MS (3, 10)	61
Fibrinogen beta chain	P02675	AD, PD (10)	56
GAP43 (neuromodulin)	P17677	AD (36)	39
Vitamin D-binding protein	P02774	AD, PD, MS (10)	38
Gelsolin	P06396	AD, MS, PD (10)	37
Insulin-like growth factor-binding protein 5	P24593	AD, PD (10)	37
Zinc-alpha-2-glycoprotein	P25311	AD, PD, MS (10)	37
Alpha-1B-glycoprotein	P04217	AD, MS, PD (10)	31
Beta-2-glycoprotein 1	P02749	AD, PD (10)	31
Actin, cytoplasmic 1	P60709	AD, MS (10)	23
Apolipoprotein D	P05090	AD, MS (3, 10)	23
Alpha-1-Antitrypsin	P01009	AD, MS (10)	22
Contactin-1	Q12860	AD, MS (10)	22
Alpha-1-acid glycoprotein 1 (Orosomucoid 1)	P02763	AD, MS, PD (10)	21
Retinol-binding protein 4	P02753	AD, MS, PD (10)	21
Polymeric immunoglobulin receptor	P01833	AD, PD (10)	18
Kallikrein-6	Q92876	AD, MS, PD (10)	17
Serum amyloid A protein precursor	P0DJI8	AD, PD (10)	17
Somatostatin	P61278	AD, Depression (3)	17

EGF-containing fibulin-like extracellular matrix protein 1	Q12805	AD, MS (10)	15
Cell growth regulator with EF-hand domain protein 1	Q99674	AD, PD (10)	14
Dystroglycan	Q14118	AD, MS (10)	14
Angiotensinogen	P01019	AD, MS, PD (10)	13
IL-6 soluble receptor	P40189	AD, Stroke (3)	13
Adherens junction-associated protein 1	Q9UKB5	AD, PD (10)	12
Ceruloplasmin	P00450	AD, MS, PD (10)	11
Neurofilament medium polypeptid	P07197	MS, MSA (3)	10
72 kDa type IV collagenase	P08253	MS, PD (10)	9
Insulin-like growth factor binding protein-3	P17936	CNS tumours (3)	8
Hypocretin-1 (Orexin-A)	O43612	Narcolepsy (3)	7
Synaptotagmin-1	P21579	AD (3)	7
Creatine kinase B-type	P12277	CJD (3)	6
Microtubule-associated protein tau	P10636	AD, CJD, PD (3)	5
Myelin basic protein	P02686	MS (3)	5
Neurexin-1-alpha	Q9ULB1	AD, PD (10)	5
Pigment epithelium-derived factor	P36955	AD, ALS, MS, PD (3, 10)	5
Pro-opiomelanocortin	P01189	Infantile Autism (3)	5
Alpha-2-macroglobulin	P01023	MS, PD (10)	4
Disintegrin and metalloproteinase domain-containing protein 17	P78536	AD (3)	4
132 kDa protein (Transcriptional-regulating factor 1)	Q96PN7	AD, PD (10)	3
Acyl-CoA-binding protein	P07108	AD (3)	3
Neurospecific enolase	P09104	CJD (36)	3
14-3-3 protein	P31947	CJD (3)	2
BACE2	Q9Y5Z0	AD (37)	2
Neurogranin	Q92686	AD (37)	2
Cochlin	O43405	AD, PD (10)	1
Glutamine Synthetase	P15104	AD (3)	1
Neurofilament heavy polypeptide	P12036	MS, MSA (3)	1
Neurofilament light polypeptide	P07196	MS; MSA (36)	1
Visinin-like protein 1	P62760	AD (37)	1

Table 4. Endogenous peptides identified from microtubule-associated protein Tau. The amino acid sequence position corresponds to isoform Tau-F (Protein accession nr. P10636-8).

Position	Sequence	Charge	MW (monoisotopic) [Da]
171-180	IPAKTPPAPK	2	1018.62
175-189	TPPAPKTPPSSGEPP	2	1458.74
175-190	TPPAPKTPPSSGEPPK	2/3	1586.83
209-224	SRTPSLPTPPTREPK	3	1662.90
212-224	TPSLPTPPTREPK	3	1419.77

Figure legends

Figure 1.

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

Figure 2.

Comparison of number of identified endogenous peptides with Mascot, Sequest HT, and PEAKS Studio.

Figure 3.

Endogenous peptides identified from microtubule-associated protein Tau (Accession no. P10636-8). The sequence of isoform Tau-F (441 aa) is shown.

Figure 4.

Identification verification of endogenous Tau 175-190 by spiking CSF with synthetic heavy isotope labelled peptide $_{175}$NH2-TPPAPKTPSSGEPPK($13C_6$)-COOH$_{190}$ (+6 Da). (A) XIC of heavy labelled (top) and endogenous (bottom) peptides shows co-elution. (B) The endogenous and heavy labelled peptide signals ($M+2H^{2+}$) are separated by 3 m/z, as expected. (C) The endogenous (bottom) and heavy (top) labelled peptides display similar fragment ion patterns.

Figure 1

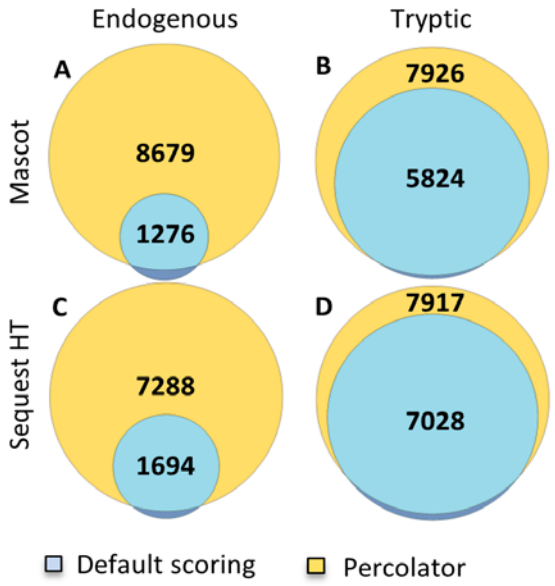

Endogenous
Tryptic
A
B
C
D
Mascot
Sequest HT
8679
1276
7926
5824
7288
1694
7917
7028
Default scoring
Percolator

Figure 2

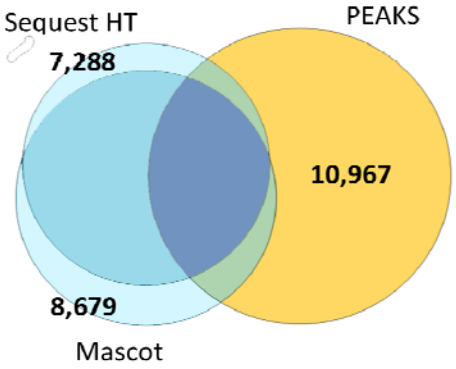


Figure 3

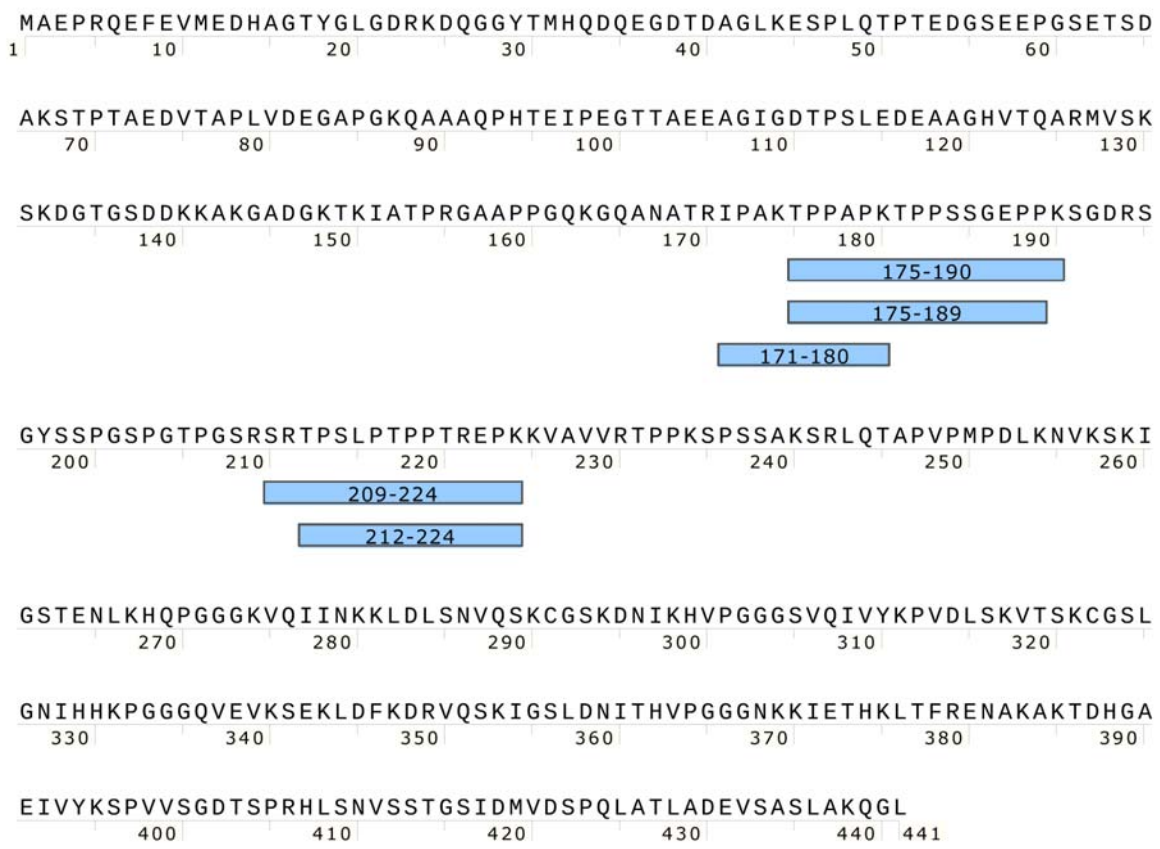
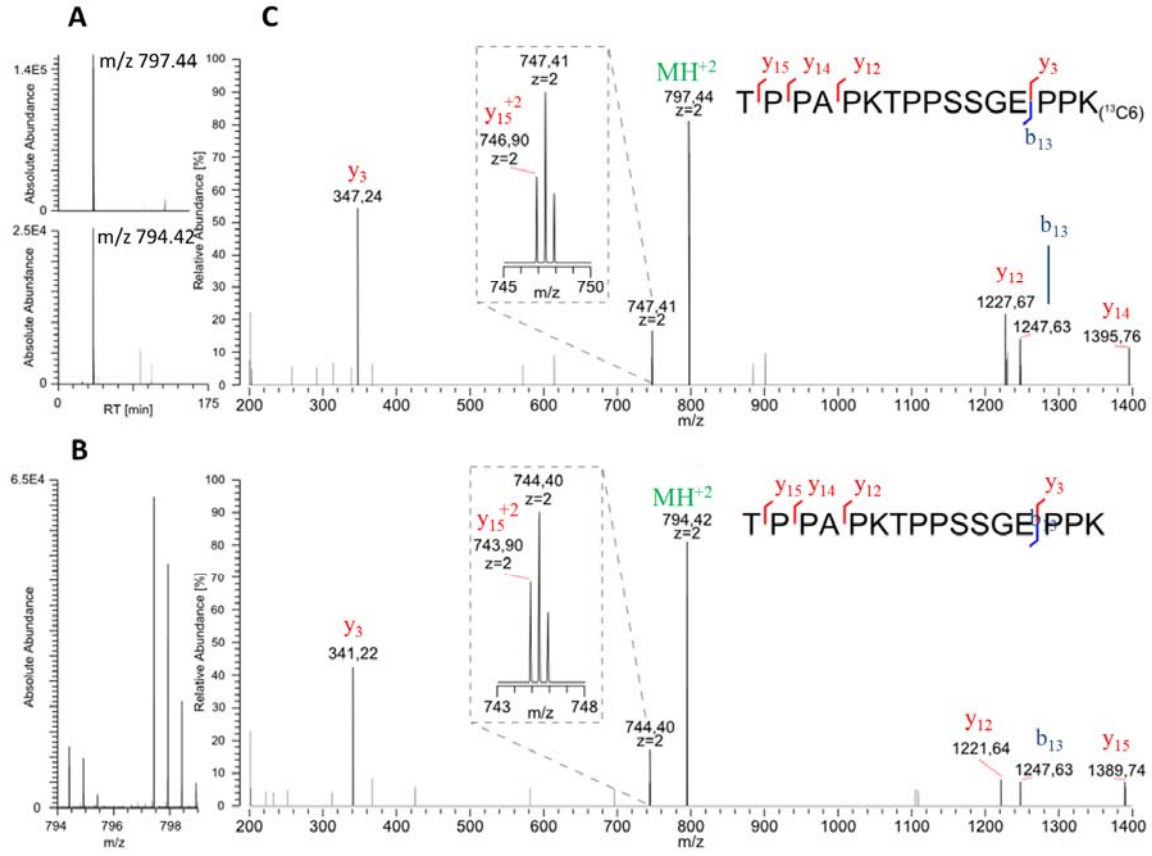

MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEEPGSETSD
1 10 20 30 40 50 60
AKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTPSLEDEAAGHVTQARMVSK
70 80 90 100 110 120 130
SKDGTGSDDKKAKGADGKTKIATPRGAAPPGQKGQANATRIPAKTPPAPKTPPSSGEPPKSGDRS
140 150 160 170 180 190
175-190
175-189
171-180
GYSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAKSRLQTAPVPMPDLKNVKSKI
200 210 220 230 240 250 260
209-224
212-224
GSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSL
270 280 290 300 310 320
GNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGA
330 340 350 360 370 380 390
EIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSASLAKQGL
400 410 420 430 440 441

Figure 4


A
C
B
m/z 797.44
m/z 794.42
Absolute Abundance
RT [min]
Relative Abundance [%]
m/z
y3
347,24
747,41
z=2
y15+2
746,90
z=2
MH+2
797,44
z=2
y15 y14 y12
y3
TPPAPKTPPSSGEPPK(13C6)
b13
b13
y12
1227,67
1247,63
y14
1395,76
744,40
z=2
y15+2
743,90
z=2
MH+2
794,42
z=2
y15 y14 y12
y3
TPPAPKTPPSSGEPPK
b13
y3
341,22
744,40
z=2
y12
1221,64
b13
1247,63
y15
1389,74