Data in Brief 32 (2020) 106048

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Human cerebrospinal fluid data for use as spectral library, for biomarker research

Lukas M. Schilde[a,b], Simone Steinbach[a,b], Bettina Serschnitzki[a,b], Fabian Maass[c], Mathias Bähr[c], Paul Lingor[c,d,e], Katrin Marcus[a,b,*], Caroline May[a,b,*]

[a] Ruhr-University Bochum, Center for Protein Diagnostics (PRODI), Medical Proteome Analysis, Germany
[b] Ruhr-University Bochum, Medical Faculty, Medizinisches Proteom-Center, Germany
[c] Department of Neurology, University Medical Center Göttingen, Göttingen, Germany
[d] Center for Biostructural Imaging of Neurodegeneration (BIN), University of Göttingen Medical Center, Göttingen, Germany
[e] Technical University of Munich, School of Medicine, Klinikum rechts der Isar, Department of Neurology, München, Germany

ARTICLE INFO

Article history:
Received 2 June 2020
Accepted 15 July 2020
Available online 19 July 2020

Keywords:
Proteomics
Mass spectrometry
Data-independent acquisition
Parkinson's disease
Cerebrospinal fluid
Spectral library

ABSTRACT

Spectral libraries generated by data dependent acquisition (DDA) are a useful tool for the analysis of data created by data independent acquisition (DIA) in mass spectrometry. The quality of DIA analysis is dependent on the quality of the spectral library. We used cerebrospinal fluid (CSF) of patients with Parkinson's disease and healthy controls to create a spectral library of human CSF proteome. To this date, there is no validated CSF biomarker for Parkinson's disease. This data set may therefore be valuable for the future analysis of CSF proteins. Part of the samples consisted of fractions that were separated by gel electrophoresis. After tryptic digestion, all samples were spiked with indexed retention time (iRT) peptides and were measured using a DDA mass spectrometry approach. The here provided data set can be used as a CSF-specific spectral library. Data files generated from the described workflow are hosted in the public repository ProteomeXchange under the identifier PXD013487.

* Corresponding authors.
E-mail addresses: Katrin.Marcus@rub.de (K. Marcus), Caroline.May@rub.de (C. May).

https://doi.org/10.1016/j.dib.2020.106048

Specifications table

Subject	Proteomics
Specific subject area	DIA based mass spectrometric analysis of human CSF
Type of data	Mass spectrometric raw data (DDA), spectral library
How data were acquired	Mass spectrometry (Q Exactive mass spectrometer operated in DDA mode performing HCD fragmentation)
Data format	Raw Unfiltered, filtered
Parameters for data collection	Intensity, retention time and charge of fragmented tryptic peptides were obtained by mass spectrometric DDA measurements of human CSF.
Description of data collection	The 127 CSF samples were collected, pooled and fractionated. Pooled CSF was fractionated by gel electrophoresis resulting in 15 bands and 5 additional technical replicates. All samples were digested with Trypsin. The samples were measured using HPLC-mass spectrometry in DDA mode and processed using Spectronaut™ Pulsar software (Biognosys AG).
Data source location	Ruhr-University Bochum, Center for Protein Diagnostics (PRODI), Medical Proteome Analysis, Bochum, Germany (51°26′43.4″N 7°15′27.9″E)
Data accessibility	Repository name: ProteomeXchange Data identification number: PXD013487 Direct URL to data: Username: reviewer16232@ebi.ac.uk Password: Zk6mtuz6
Related research article	K. Barkovits, A. Linden, S. Galozzi, L. Schilde, S. Pacharra, B. Mollenhauer, N. Stoepel, S. Steinbach, C. May, J. Uszkoreit, M. Eisenacher, K. Marcus, Characterization of Cerebrospinal Fluid by Data-Independent Acquisition Mass Spectrometry, J Proteome Res. 2018. https://doi.org/10.1021/acs.jproteome.8b00308 [1]

Value of the data

- The data set can be used as a spectral library for CSF in the context of neurodegenerative disorders, especially Parkinson's disease
- It may assist in the search for biomarkers in neurodegenerative diseases
- It may help optimize parameters for identification of proteins/peptides and thus expand the proteome of cerebrospinal fluid
- It may help to uncover protein pathways in CSF, that originate in the brain

1. Data description

The here presented proteomic dataset offers mass spectrometry data files generated from human CSF. The CSF was collected by lumbar puncture. Afterwards, proteins were eluted, the protein concentration determined and proteins separated by gel electrophoresis. The resulting lanes were cut into single bands followed by an in-gel digestion with trypsin. Additionally, five pooled samples were digested using an in-solution approach. iRT peptides were added to each sample. All samples were measured with a data-dependent based mass spectrometric approach. All samples are provided in the raw-file and mzML-format. All samples were used to create an exemplary spectral library in Spectronaut™ Pulsar software (Biognosys AG) and it was provided alongside the raw data.

2. Experimental design, materials and methods

2.1. Sampling

CSF sample collection was performed referring to standardized biomarker guidelines [2]. CSF was centrifuged, aliquoted and stored at −80 °C in polypropylene tubes within 1–2 h after lumbar puncture. Samples with relevant blood contamination (erythrocyte count > 100 /μL) were excluded from further analysis.

2.2. Amino acid analysis

The protein concentration of the CSF was measured by amino acid analysis [3]. To avoid contaminations, glass vials were incubated in a muffle furnace (muffle furnace, Carbolite CWF 1100) at 400 °C for four hours. In each clean glass vial, 4 µL of the CSF was transferred, dried in a vacuum concentrator (RVC2–25CD plus, Martin Christ Gefriertrocknungsanlagen) and placed in an evacuation vessel. For each sample, two vials were prepared. Then, 400 µL 6 M hydrochloric acid and a phenol crystal were added. The samples were alternately evacuated four times and aerated with argon. The peptides were then hydrolyzed into amino acids by acidic gas phase hydrolysis. The remaining hydrochloric acid was removed by a new evacuation. Then 10 μL of 20 mM hydrochloric acid were added to the samples. In the next step, 30 μL AccQ fluoroborate buffers with the internal standard Norleucine and 10 μL AccQ fluoror reagent (10 mM 6-aminoquinolyl-N-hydroxysuccinimidylcarbamate in acetonitrile) were added to derivatize the amino acids. After that, the samples were incubated at 55 °C for 10 min. The resulting derivatives were loaded and separated on a C18 reversed-phase separation column (2.1 mm x 100 mm length, Waters GmbH). Two solvents (solvent A: AccQ-Tag Ultra Eluent A, solvent B: AccQ-Tag Ultra Eluent B) were used in the gradient system. The elution of the derivatives was performed with a flow rate of 0.7 mL per minute and a column temperature of 55 °C. The detection of the amino acid derivatives was performed by UV spectrophotometry (Waters GmbH) at an emission wavelength of 260 nm. Different concentrations of an internal amino acid standard were used to allow quantification. The protein concentration of the CSF samples could then be determined by considering the molar mass of the amino acids and the volume of the CSF samples.

2.3. Sample preparation and gel electrophoresis

Human CSF of 127 individuals was pooled, lyophilized (Alpha 2–4 LDplus, Martin Christ Gefriertrocknungsanlagen GmbH) and resuspended in 50 µL distilled water. After this, gel electrophoresis was performed. In total 80 µg precipitated proteins were loaded on a NuPAGE™ 3–8% Tris-Acetate gel (both from Fisher Scientific GmbH). Additionally, 20 µg precipitated proteins were loaded on a NuPAGE™ 4–12% BisTris gel (both from Thermo Fisher Scientific). Gel electrophoresis was performed at 200 mA and 200 W, with the following voltage profile: 50 V for 15 min and 180 V for 50 min. Subsequently, gels were stained with Coomassie Blue Staining (SimpleBlue™ SafeStain, Thermo Fisher Scientific) according to the manufacturers' protocol.

2.4. Destaining and in-gel-digestion

Protein lanes were cut into single fractions (in total 10 for TrisAcetate and 6 for BisTris), destained and pH adjusted by incubating the gel pieces for 10 min with 50 mM ammonium bicarbonate (Sigma-Aldrich) as well as 50% (v/v) 50 mM ammonium bicarbonate with 50% (v/v) and 100% acetonitrile (Merck KGaA) three times alternately. After the first incubation with ammonium bicarbonate and acetonitrile, samples were treated with 50 µL 10 mM 1,4-dithiothreitol

(DTT) (AppliChem GmbH) for 1 h at 56 °C. In the next step, 50 µL 50 mM iodoacetamide (Merck KGaA) were incubated for 45 min at room temperature in the dark. Finally, gel pieces were dried in a vacuum concentrator (RVC2–25CD plus, Martin Christ Gefriertrocknungsanlagen) and resuspended in 6 µL 50 mM ammonium bicarbonate. Overnight digestion was initiated by adding 6 µL of trypsin solution (0.012 µg/µL, Promega Corp.). The digestion was stopped and peptides eluted by incubating the gel pieces two times for 15 min with 20 µL of a 1:1 solution containing 100% acetonitrile and 0.1% (v/v) trifluoroacetic acid (TFA) (Merck KGaA) in an ice-cooled ultrasonic bath. Samples were transferred in glass vials, dried in a vacuum concentrator and resuspended in 20 µL 0.1% (v/v) TFA. The protein concentration was measured by using amino acid analysis, for this purpose, a double determination with 4 µL was performed [4]. Up to 300 µg were transferred into glass vials. Finally, samples were injected with 1 µL of iRT peptides (Biognosys AG).

2.5. In-solution-digestion

2 μL of 0.2% (w/v) RapiGest^{TM} SF Surfactant (dissolved in 50 mM ammonium bicarbonate) were added to each sample. Subsequently, 0.8 μL of 250 mM DTT was added to the samples. Incubation was performed for 30 minutes at 60 °C. After cooling down to room temperature, samples were incubated in the dark with 1.12 μL of 0.55 M iodoacetamide for 30 min. Proteolytic digestion was initiated by addition of 2 μL of trypsin solution (0.1 μg/μL trypsin in 50 mM ammonium bicarbonate and 3 mM acetic acid) and was performed for 16 h at 37 °C and 350 rpm, overnight.

To stop digestion, 2,3 μL 10% (v/v) TFA were applied to each sample. After incubation for 45 min at 37 °C and 350 rpm, samples were centrifuged for 10 min at 4 °C and 20,817 G (Centrifuge 5417R, Eppendorf AG). Supernatants were transferred into glass vials and were dried using a vacuum concentrator. Finally, peptides were resuspended in 20 μL 0.1% (v/v) TFA and stored at −80 °C. For each sample 8 μL was taken and protein concentration measured with amino acid analysis (Plum et al.). 2 µL of each sample were pooled, creating a master mix.

The samples were spiked with 1 µL injection volume of iRT peptides (Biognosys AG) before they were analyzed by HPLC-MS/MS.

2.6. High-performance liquid chromatography

The extracted peptides were concentrated and cleaned-up by using the nano-HPLC system Ultimate 3000 (Dionex, Thermo Fisher Scientific GmbH). It contained both an integrated capillary pre-column (100 µm × 2 cm, particle size 5 µm, pore size 100 Å; Thermo Scientific Corp.), as well as an analytical column (PepMap C18 75 µm x 50 cm, particle size 2 µm, pore size 100 Å; Thermo Scientific Corp.). The precolumn was washed for 7 min with 0.1% TFA. The peptides were eluted from the pre-column to the analytical column. The peptides were separated by applying a stepwise 2 h and 15 min gradient of buffer A (0.1% fluoric acid) and buffer B (84% acetonitrile, 0.1% fluoric acid) with a flow rate of 400 nL•min^{-1}.

The gradient was run from 5 to 41% buffer for 117 min and from 41% to 52% for 3 min followed by a 5 min washing step at 95% buffer B and a 5 min equilibration step at 5% buffer B. The column oven temperature was set to 60 °C.

2.7. Mass spectrometry settings

The peptides were ionized by electrospray ionization (ESI) and injected into a Q Exactive mass spectrometer (Thermo Fisher Scientific GmbH). The instrument operated in DDA mode. Thus, it performed HCD fragmentation of the top 10 abundant precursor ions at stepped NCE

25.5, 27 and 30. The mass range was set to 400–1400 *m/z* with a resolution of 70,000 at 200 *m/z* (AGC 3e6, 80 ms maximum injection time). The capillary temperature was set to 250 °C and the spray voltage to 1600 V. The fragment analysis was performed in an orbitrap mass analyzer with a resolution of 35,000 at 130 *m/z* (AGC 2e5, 120 ms maximum injection time, 2.0 *m/z* isolation width, loop count: 10).

The spectral library was generated in the interface of the Spectronaut™ Pulsar software (Biognosys AG). Therefore, generated spectra of DDA measurements were aligned with theoretical spectra present in the reference database of Uniprot/SwissProt human proteome database (UniProtKB/Swiss-Prot UniProt release 2017_01; downloaded 2017–01–26; Number of entries 553,474). In total 20 DDA (15 fractions and 5 technical replicates of pooled samples) measurements were used. Default settings of Spectronaut™ Pulsar were applied for the generation of the spectral library. The false discovery rate (FDR), as well as the Q-value, had a cutoff-value of 1%. The generated spectral library was exported in the excel-format.

The resulting raw-files were converted into the mzML-format and uploaded alongside the excel-file in the public repository PRIDE ProteomeXchange under the identifier PXD013487.

Ethics statement

Informed consent of each subject was obtained before CSF sampling.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships, which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

This work was supported by the Deutsche Parkinson Gesellschaft, Medical Faculty at RUB (FoRUM), the European Union (SILENCE, FKZ 01EW1701A; NISCI, GA no. 681094), the German Federal Ministry of Education and Research (WTZ with Brazil, FKZ 01DN14023), the HUPO Brain Proteome Project (HBPP) and PURE, a project of North Rhine-Westphalia, a federal German state.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.106048.

References

[1] K. Barkovits, et al., Characterization of Cerebrospinal fluid via data-independent acquisition, mass spectrometry, J. Proteome. Res. 17 (10) (2018) https://doi.org/10.1021/acs.jproteome.8b00308.
[2] C.E. Teunissen, et al., A Consensus Protocol for the Standardization of Cerebrospinal Fluid Collection and Biobanking, Neurology 73 (22) (2009) https://doi.org/10.1212/WNL.0b013e3181c47cc2.
[3] Annika Guntermann, et al., Human Tear Fluid Proteome Dataset for Usage as a Spectral Library and for Protein Modeling, Data Brief. 23 (2019) https://doi.org/10.1016/j.dib.2019.103742.
[4] S. Plum, et al., Proteomic Characterization of Neuromelanin Granules Isolated from Human Substantia Nigra by Laser-Microdissection, Sci. Rep. 6 (2016) https://doi.org/10.1038/srep37139.