Proteome Preparation and Analysis

Bryan Yoo[1], Jessica Griffiths[1], Sarkis Mazmanian[1]

[1]California Institute of Technology

Nov 02, 2021

1

dx.doi.org/10.17504/protocols.io.bzqcp5sw

Mazmanian Lab

Jessica Griffiths

Protocol for proteome preparation and analysis used in Yoo et al 2021

DOI

dx.doi.org/10.17504/protocols.io.bzqcp5sw

Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian 2021. Proteome Preparation and Analysis. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bzqcp5sw

Center for Environmental and Microbial Interactions (CEMI)
Grant ID: n/a

Emerald Foundation
Grant ID: n/a

Heritage Medical Research Institute
Grant ID: n/a

National Institutes of Health
Grant ID: GM007616 and DK078938

Department of Defense
Grant ID: PD160030

Aligning Science Across Parkinson's
Grant ID: ASAP-000375

protocol ,

Nov 02, 2021

Nov 02, 2021

Protein Extraction

1 Mice were sacrificed 45 minutes after C21 administration and cecal contents were isolated and resuspended in 400µl of phosphate buffered solution, and centrifuged at x20,000g to spin down cells and lysate.

2 Protein was isolated from resulting supernatant using Wessel-Flüegge's methanol/chloroform extraction method (**Wessel and Flügge, 1984**).

3 Briefly, MeOH and chloroform was added to sample at a 4:1 and 1:1 ratio, respectively. Next, water dH2O was added at a 3:1 ratio, samples were vortexed and centrifuged at x20,000g.

4 Resulting precipitated protein was collected and washed with methanol. Precipitated protein was centrifuged and was left to air dry, and stored in −20C until protein digestion.

In-solution Protein Digestion and Desalting

5 Precipitated protein sample was denatured in 40µl of 8M Urea (100mM Tris-HCl pH8.5).

6 To reduce disulfide bonds, 1.25µl of 100mM Tris(2-carboxyethyl)Phosphine was added and incubated at room temperature (RT) for 20 minutes.

7 Then 1.8µl of 250mM iodoactaminde was added and incubated at RT in the dark to alkylate cysteines.

8 The first step of digestion was initiated by adding 1µl of 0.1µg/µl of Lysyl endopeptidase. After four-hour incubation. Urea concentration was adjusted to 2M by adding 120µl of 100mM Tris-HCl pH8.5.

9 The second step of digestion was initiated by adding 2.5µl of 2µg/µl trypsin plus 1.6µl of 100mM CaCl2 for trypsin activity enhancement and incubating overnight in the dark. Formic acid was added to stop trypsing digestion.

The digested peptides were desalted by HPLC using C8 peptide microtrap (Microm

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Proteome Preparation and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqcp5sw

10 Bioresources). Peptides were lyophilized and diluted to 200 ng/µl in 0.2% formic acid prior to LC-MS/MS analysis.

LC-MS/MS

11 Samples were analyzed on a Q Exactive HF Orbitrap mass spectrometer coupled to an EASY nLC 1200 liquid chromatographic system (Thermo Scientific, San Jose, CA). Approximately 200 ng of peptides were loaded on a 50 µ I.D. × 25 cm column with a 10 µ electrospray tip (PicoFrit from New Objective, Woburn, MA) in-house-packed with ReproSil-Pur C18-AQ 1.9 µ (Dr. Maisch, Ammerbuch, Germany). Solvent A consisted of 2% MeCN in 0.2% FA and solvent B consisted of 80% MeCN in 0.2% FA. A non-linear 60 minute gradient from 2% B to 40% B was used to separate the peptides for analysis. The mass spectrometer was operated in a data-dependent mode, with MS1 scans collected from 400-1650 m/z at 60,000 resolution and MS/MS scans collected from 200-2000 m/z at 30,000 resolution. Dynamic exclusion of 45 s was used. The top 12 most abundant peptides with charge states between 2 and 5 were selected for fragmentation with normalized collision energy of 28.

Peptide and Protein Identification

12 Thermo .raw files were converted to .ms1 and .ms2 files using RawConverter 1.1.0.18 (**He et al., 2015**) operating in data dependent mode and selecting for monoisotopic m/z.

13 Tandem mass spectra (.ms2 files) were identified by database search method using the Integrated Proteomics Pipeline 6.5.4 (IP2, Integrated Proteomics Applications, Inc.,http://www.integratedproteomics.com).

14 Briefly, databases containing forward and reverse (decoy) (**Elias and Gygi, 2007**;**Peng et al., 2003**) peptide sequences were generated from in silico trypsin digestion of either the mouse proteome (UniProt; Oct. 2, 2019) or protein sequences derived from large comprehensive public repositories (ComPIL 2.0) (**Park et al., 2018**).

15 Tandem mass specta were matched to peptide sequences using the ProLuCID/SEQUEST (1.4) (**Xu et al., 2015**;**2006**) software package. The validity of spectrum-peptide matches were assessed using the SEQUEST-defined parameters XCorr (cross-correlation score) and DeltaCN (normalized difference in cross-correlation scores) in the DTASelect2 (2.1.4) (**Cociorva et al., 2007**;**Tabb et al., 2002**) software package. Search settings were configured as follows: (1) 5ppm precursor ion mass tolerance, (2) 10ppm fragment ion mass tolerance, (3) 1% peptide false discovery rate, (4) 2 peptide per protein minimum, (5) 600-6000Da precursor mass window, (6) 2 differential modifications per peptide maximum (methionine oxidation: M+15.994915 Da), (7) unlimited static modications per peptide (cysteine carbamidomethylation: C+57.02146 Da), and (8) the search space included half- and fully tryptic (cleavage C-terminal to K and R residues) peptide candidates with unlimited (mouse database, custom metagenomic shotgun database) or 2 missed cleavage events (ComPIL 2.0).

Differential Analysis of Detected Proteins using Peptide-Spectrum Matches (Spectral Counts)

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Proteome Preparation and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqcp5sw

16 Detected proteins were grouped by sequence similarity into “clusters” using CD-HIT 4.8.1 (**Fu et al., 2012**;**Li and Godzik, 2006**;**Li et al., 2001**) at the following similarity cut-offs: 65%, 75%, 85%, and 95%. The following is an example command line input: “cd-hit -i fastafile.fasta -o outputfile -c 0.65 -g 1 -d 0”.

17 Tandem mass spectra identified as peptides (peptide spectrum matches, PSMs) were mapped to CD-HIT generated clusters. PSMs mapping to >1 cluster were discarded. Cluster-PSM tables were generated and differential analysis was performed in DESeq2 (1.25.13) (**Love et al., 2014**).

18 Briefly, count data (PSMs) were modeled using the negative binomial distribution, and the mean-variance relationship was estimated. Variance was estimated using an information sharing approach whereby a single feature’s (or cluster’s) variance was estimated by taking into account variances of other clusters measured in the same experiment.

19 Feature significance calling and ranking were performed using estimated effect sizes. Multiple testing correction was performed by Benjamini-Hochberg method within the DESeq2 package. Volcano plots were generated in Prism (GraphPad).

Differential Analysis of Detected Proteins using Ion Intensity (Precursor Intensity)

20 Detected proteins were grouped into “clusters” by sequence similarity using CD-HIT 4.8.1 (**Fu et al., 2012**;**Li and Godzik, 2006**;**Li et al., 2001**) at the following similarity cut-offs: 65%, 75%, 85%, and 95%. The following is an example command line input: “cd-hit -i fastafile.fasta -o outputfile -c 0.65 -g 1 -d 0”.

21 Using the Census software package (Park et al.; 2008) (Integrated Proteomics Pipeline 6.5.4), peptide ion intensities were calculated from .ms1 files. Peptide ion intensities were assigned to their parent peptide, then parent peptides were mapped to their appropriate CD-HIT generated clusters. Ion intensities belonging to parent peptides that map to >1 CD-HIT cluster were discarded. Cluster-ion intensity tables were generated.

22 Ion intensity data were analyzed using the DEP package (**Zhang et al., 2018**) operating in R. Intensity values were automatically Log2 transformed in DEP. The cluster list was subsequently filtered with the ‘filter_proteins’ function such that clusters with missing values above a 65% threshold were discarded.

23 Remaining intensities were further transformed by the ‘normalize_vsn’ function (**Huber et al., 2002**). Missing data in remaining clusters were imputed using a mixed approach. Clusters where either the control or treatment group contained only null entries were classified as ‘missing not at random’ (MNAR) and imputed with 0 values. All other groups were treated as ‘missing at random’ (MAR) and imputed using the maximum likelihood method (‘MLE’) (**Gatto and Lilley, 2012**). Note that for a given cluster, missing values for treatment groups were imputed separately by treatment group.

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Proteome Preparation and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqcp5sw

24 Differential expression analyses were performed on filled-in cluster-ion intensity tables using the 'test_diff' function (**Ritchie et al., 2015**) and multiple testing correction was performed using the 'add_rejections' function.

Network Analysis using STRING Database

25 Upregulated proteins with a nominal p-value<0.2 were searched against protein-protein interactions in the STRING database (http://www.string-db.org) where high confidence interactions were selected for. Briefly, the STRING database sources protein-protein interactions from primary databases consisting of genomic context predictions, high-throughput lab experiments, (conserved) co-expression, automated textmining, and pervious knowledge in databases (**Szklarczyk et al., 2019**).

Metaproteome Analysis using Unipept

26 Upregulated tryptic, microbial peptide sequences with nominal p-value cutoff of p<0.2 were input into Unipept (http://unipept.ugent.be) (**Gurdeep Singh et al., 2019**;**Mesuere et al., 2015**), equating leucine and isoleucine and filtering duplicate peptides. Briefly, Unipept indexes tryptic peptide sequences from the UniProtKB and details peptides with NCBI's taxonomic database. Lowest common ancestor is calculated for each tryptic peptide.

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Proteome Preparation and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqcp5sw