Data from quantitative label free proteomics analysis of rat spleen

Data in Brief 8 (2016) 494–500

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Data from quantitative label free proteomics analysis of rat spleen

Khadar Dudekula, Thierry Le Bihan *

SynthSys, University of Edinburgh, Waddington Building, The King's Buildings, Max Born Crescent, Edinburgh EH9 3BF, United Kingdom

ARTICLE INFO

Article history:
Received 2 March 2016
Received in revised form
17 May 2016
Accepted 24 May 2016
Available online 4 June 2016

Keywords:
Spleen
Rat
Protein extraction
Label-free quantitative proteomics

ABSTRACT

The dataset presented in this work has been obtained using a label-free quantitative proteomic analysis of rat spleen. A robust method for extraction of proteins from rat spleen tissue and LC-MS-MS analysis was developed using a urea and SDS-based buffer. Different fractionation methods were compared. A total of 3484 different proteins were identified from the pool of all experiments run in this study (a total of 2460 proteins with at least two peptides). A total of 1822 proteins were identified from nine non-fractionated pulse gels, 2288 proteins and 2864 proteins were identified by SDS-PAGE fractionation into three and five fractions respectively. The proteomics data are deposited in ProteomeXchange Consortium via PRIDE PXD003520, Progenesis and Maxquant output are presented in the supported information. The generated list of proteins under different regimes of fractionation allow assessing the nature of the identified proteins; variability in the quantitative analysis associated with the different sampling strategy and allow defining a proper number of replicates for future quantitative analysis.

* Corresponding author. Tel.: +44 131 651 9073.
E-mail address: thierry.lebihan@ed.ac.uk (T. Le Bihan).

http://dx.doi.org/10.1016/j.dib.2016.05.058

Specifications Table

Subject area	Biology
More specific subject area	Protein extraction from rat tissues, quantitative proteomics
Type of data	Figures, Excel sheets
How the data was acquired	In-gel trypsin digestion of SDS-PAGE gels followed by LC-MS analysis in data-dependent mode using a Q-Exactive hybrid quadrupole-orbitrap mass spectrometer. (LC:Ultimate3000 HPLC system,Thermo-Fisher Scientific. MS: Orbitrap QExactive,Thermo-FisherScientific)
Data format	Raw, processed
Experimental factors	Protein extract from rat spleen samples were run on SDS-PAGE for 5 min (pulse gel, pg) or full length separated. The gels were cut into 3 to 5 fractions each followed by in-gel digestion with trypsin.
Experimental features	Tissue samples were homogenised in 8M Urea with 1% SDS. Gel fractionation was employed before gradient separation in HPLC followed by MS analysis. Data identification performed with Mascot and Maxquant, the label-free quantitation was performed with Progenesis.
Data source location	Centre for Synthetic and Systems Biology, CH. Waddington Building, The University of Edinburgh, Max Born Crescent, Kings Buildings, Edinburgh, EH9 3BF. United Kingdom.
Data accessibility	Data are available within this article and from ProteomeXchange Consortium Via PRIDE partner repository PRIDE: PXD003520.

Value of the data

- Optimised sample preparation and protein extraction from rat spleen tissue using SDS/urea based buffer system (8 M urea and 1% SDS).
- A robust MS/MS method for quantitative proteomics analysis of the rat spleen proteome enabling to identify *ca*. 1800 proteins without fractionation and up to 3400 proteins combining all the different fractionations strategies.
- Generation of a list of proteins with and without fractionation, under different conditions of replication which allows defining the number of replicates. The provided lists of identified proteins under different regimes of fractionations is a useful tool for follow-up experiments for any researcher working in the field of immunology and tissue analysis. A clear yield in protein extraction per amount of tissues is as well presented.

1. Data

MS Data processed using Mascot (S2), Progenesis (S3) and Maxquant (S4) are presented in the supplementary information tables. An additional supplementary table is presented to highlight the experimental design (S1).

2. Experimental design

We present a dataset obtained from using the label free proteomics analysis of rat spleen tissue. In total four different rat spleen samples were used to generate the proteomics dataset (experimental

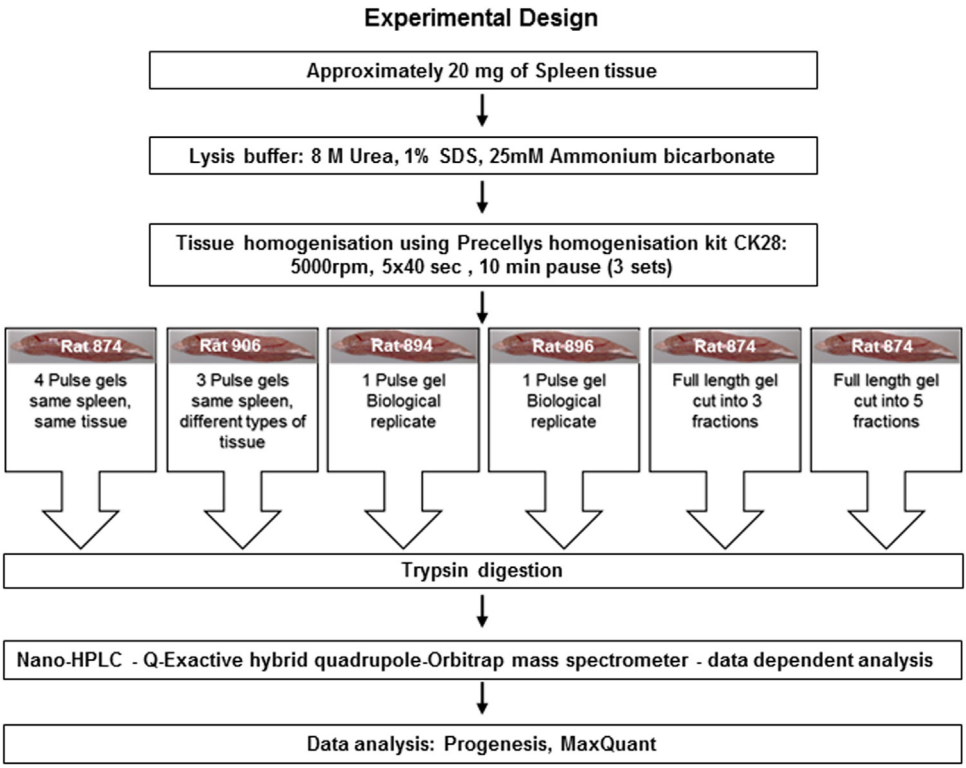

Fig. 1. Schematic overview of the overall analysis workflow. Sample preparation variability tested at the LC-MS injection and sample digestion (technical variability) and biological replicates of similar spleens were as well compared.

design shown in Fig. 1) and were analysed by resolving the proteins on SDS-PAGE either by ten minutes pulse gel (pg) as well as resolved on a full-length gel and cut into 3 and 5 gel fractions respectively. Three separate extractions were prepared from one single spleen as technical replicates and two individual extractions and combine with two different spleens (as biological replicates). The sampling and experimental design are shown in Fig. 1. The dataset was generated considering the type of variation in the tissue as

(I) Dataset of Proteins analysed from a single rat spleen tissue taken from one single extraction injected multiple times in MS (technical replicates assessment of the LC-MS variation only).
(II) Dataset of Proteins analysed from three different sampling from the same rat spleen which captures variation across the same spleen.
(III) Dataset of Proteins analysed from three different biological replicates of rat spleen samples from the preparation of three different protein extractions. These samples are variable in both the sample type and the spleen source.

3. Materials and methods

3.1. Sample materials

The rat spleens (strain: Sprague Dawley) were bought from Sera Laboratories internationals Ltd. United Kingdom. The whole spleen was removed from non-medicated and non-immunised rats, rinsed in PBS and snap frozen in isopentane on dry ice before shipping.

3.2. Protein extraction

Approximately 100 mg of rat spleen tissue was sliced and roughly homogenised by cutting on a glass plate with a scalpel. An amount of 20 mg of homogenised tissue was weighed into 2 ml tubes of Precellys Lysing Kit (CK28) containing six 2.8 mm ceramic beads and 500 μl of lysis buffer (8 M urea, 1% SDS, 25 mM ammonium bicarbonate). The tubes were chilled for 5 min on ice and tissues were lysed using Precellys Evolution homogenizer (Bertin Technologies) using the following programme which consists of the following sequence: 5000 rpm, 5 cycles for 40 s with a 10 s pause between the cycles. The sequence were repeated 3 times and the tubes were rested on ice for 2 min between each sequence. After homogenisation, the lysate was sonicated in cold water bath for 2 min and centrifuged at 2655*g* for 5 min at 4 °C. The lysate was transferred to fresh Eppendorf tubes and spun again for 5 min before determining the protein concentration using Pierce BCA protein assay Kit. For approximately 20 mg of tissue about 3 mg (approximately 2.5 μg/μl) of total protein was obtained using this method.

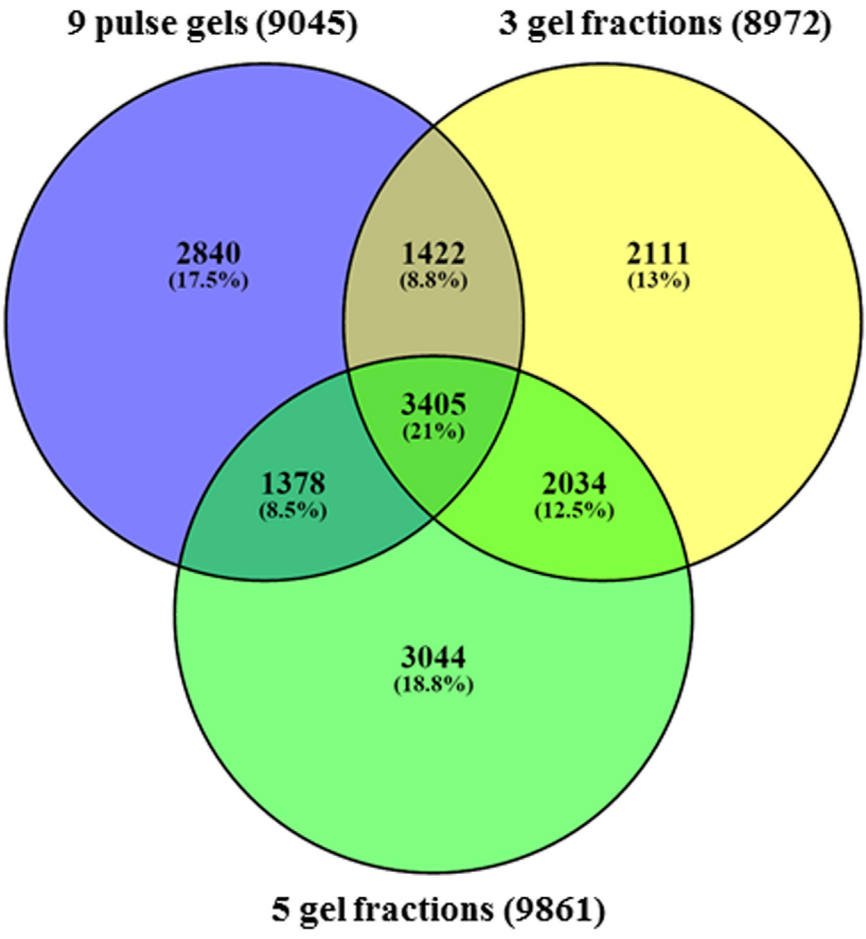

Fig. 2. Three different analysis methods were compared using a Venn diagram visualisation of peptides identified using VENNY (http://bioinfogp.cnb.csic.es/tools/venny/).

3.3. Sample preparation for MS analysis

Two types of SDS-PAGE gel fractionation methods were used in this study, either a 10 min pulse gel or a complete gel resolution was generated. An equivalent of 50 μg protein was reduced in Thermofisher NuPAGE® LDS sample buffer (4 ×) at 90 °C for 5 min. A pulse gel was run where the sample is allowed to enter the gel before any protein starts to separate which correspond to approximately 10 min up to the separation of the 4th protein marker is visible (Novex SeeBlue Plus2 Prestained standard) and two sample replicates were run to full length for gel fractionation. Two full-length lanes were cut into three fractions and five fractions respectively. The samples that were fractionated were chosen from one of the samples used for pulse gels. The gel bands were cut and trypsin digestion was performed according to [1].

3.4. HPLC-MS analysis

Peptide extracts from gel bands were then cleaned on SPE reverse phase Bond Elut LMS cartridge, 25 mg (Agilent, UK) following the manufacturer's instructions. The samples were dried under low pressure (Thermo Jouan, UK) and stored at −20 °C. The dried peptide samples were resuspended in resuspension buffer (0.5%v/v trifluoroacetic acid in water) to give final concentration of approximately 1.75 μg/μl. Before injection on LC-MS, samples were filtered using Millex 0.45 μm filter. Nano-HPLC-MS/MS analysis was performed using an on-line system consisting of a nano-pump (Dionex Ultimate 3000, Thermo-Fisher, UK) coupled to a QExactive instrument (Thermo-Fisher, UK) with a pre-column of 300 μm × 5 mm (Acclaim Pepmap, 5 μm particle size) connected to a column of 75 μm × 50 cm (Acclaim Pepmap, 3 μm particle size). Samples were analysed on a 90 min gradient in data dependent analysis (1 survey scan at 70k resolution followed by the top 10 MS/MS) a mass range from 400 to 2000 amu using a method similar to [2].

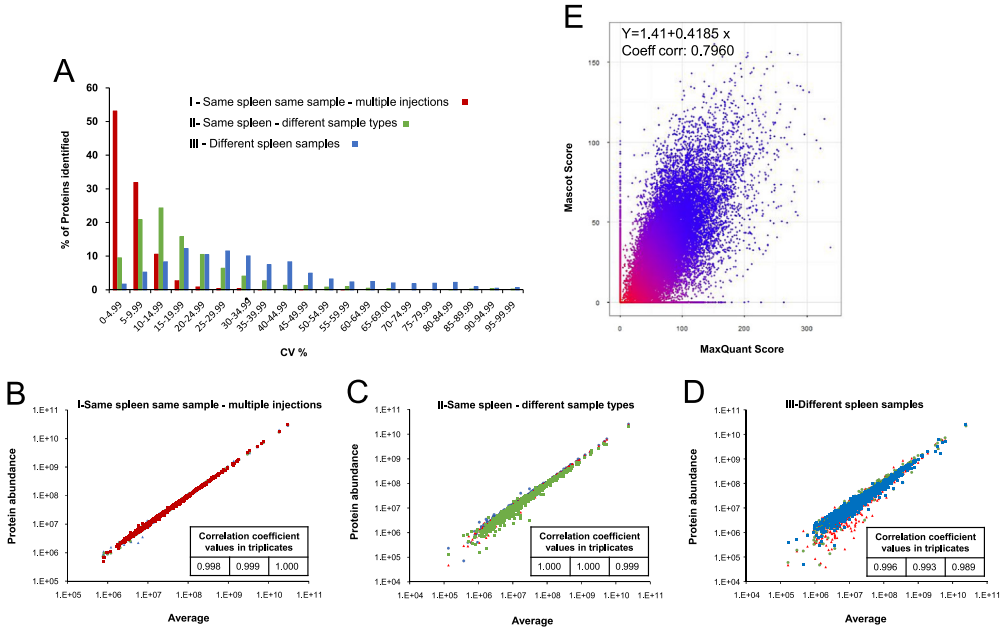

Fig. 3. (A) Coefficient variation extracted from technical LC injection (I), from a different biopsy of the same spleen (II) and different spleens of similar rats (III). (B), (C) and (D) Scattered plots of the peptide intensities extracted from the different replicates against the median group with correlation coefficient extracted for each replicate (for a Mascot Score of 20 per peptide). (E) A scatter plot of each peptide scores evaluated by Mascot or MaxQuant (no score cut-off used), the colour reflects density; red is highly dense zone and blue is less dense.

3.5. Data analysis

Data from MS/MS spectra were searched using MASCOT Versions 2.4 (Matrix Science Ltd., UK) against the Rat Uniprot Fasta dataset downloaded from 2014 April 1st with 33,892 sequences. Analysis performed with maximum missed-cut value set to 2. Following features were used in all searches: (i) variable methionine oxidation, (ii) fixed cysteine carbamidomethylation, (iii) precursor mass tolerance of 10 ppm, (iv) MS/MS tolerance of 0.05 amu, (v) significance threshold (p) below 0.05 (MudPIT scoring) and (vi) minimal peptide Mascot score of 20. In some instance data were converted using ProteoWizard MSConvert (64 bits) Version 2.1.x [3] and data was merged using Mascot Daemon 2.5.1 (Matrix Science). Dataset of individual searches are shown in Supplementary information 1.

Progenesis (version 4 Nonlinear Dynamics, UK) was used for LC-MS label-free intensity based quantitation. Only MS/MS peaks with a charge of 2+, 3+ or 4+ were taken into account for the total number of 'Features' (signal at one particular retention time and m/z) and only the five most intense spectra per 'Feature' were included. Normalisation was performed based on the median of the ion intensities of these sets of multi-charged ions (2+, 3+, and 4+) against a reference spectra. The associated unique peptide ion intensities for a specific protein were then summed to generate an abundance value, from which was then transformed using an ArcSinH function. One-way ANOVA was used to calculate the p-value based on the transformed values. Dataset of the Progenesis analysis is shown in Supplementary information 3.

The same dataset was as well processed using Maxquant version 1.5.2.8 [4] using default parameters searches was performed using Andromeda using similar search criterion as from Mascot. Dataset associated to the Maxquant analysis is shown in Supplementary information 4.

Fig. 2 was generated using data obtained from MASCOT. The figure shows a total of 9045 peptides were identified; 9863 peptides were identified from a full-length gel cut into five gel fractions and with three gel fraction, 8972 peptides were identified. 3405 peptides were common peptides from all three different type of fractionation analysis. Fig. 3A shows the coefficient of variation (CV) values calculated from the intensity-based data generated from Progenesis analysis. The figure gives the percentage of proteins distributed in the range of 5% increment in CV value. Fig. 3B–D shows the scatter plots of protein abundance of three replicates for each data set as explained in the experimental design, clearly showing more variability is observed from biological replicates. In Fig. 3E, all peptides identified in this study using the two search engine Mascot and Maxquant are plotted against each other (no score cut-off were used here, for a total of 60,304 peptides). Although a positive global trend is observed, the observed coefficient correlation was estimated at 0.7960 between the scores measured by Mascot and Maxquant.

Data availability

Data are available from ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) [5] PRIDE: PXD003520; Project Name: Rat spleen proteomics profiling Project accession: PXD003520; Project DOI: 10.6019/PXD003520; Reviewer account details: Username: reviewer67804@ebi.ac.uk; Password: tyP0Zh1t.

Funding

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 602470. The LC-MS QExactive was purchased by the Centre for Immunity, Infection and Evolution supported by a Wellcome Trust grant.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.05.077.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.05.058.

References

[1] J.V. Olsen, M. Mann, A. Shevchenko, H. Tomas, J. Havlis, In-gel digestion for mass spectrometric characterization of proteins and proteomes, Nat. Protoc. 1 (6) (2007) 2856–2860.
[2] C. Millet, D. Ausiannikava, T. Le Bihan, S. Granneman, S. Makovets, Cell populations can use aneuploidy to survive telomerase insufficiency, Nat. Commun. 6 (2015) 8664. http://dx.doi.org/10.1038/ncomms9664.
[3] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol. 30 (10) (2012) 918–920.
[4] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (12) (2008) 1367–1372.
[5] J.A. Vizcaíno, A. Csordas, N. Del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, 2016 Update of the PRIDE database and its related tools, Nucleic Acids Res. 44 (D1) (2016) D447–D456.