Data in Brief 42 (2022) 108151

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Proteomics datasets of developing rat brain: Synaptic proteome and SUMO2/3-ylome

Félicie Kieffer, Marie Pronot, Anne-Sophie Gay, Delphine Debayle, Carole Gwizdek*

Université Côte d'Azur, Centre National de la Recherche Scientifique, Institut de Pharmacologie Moléculaire et Cellulaire, France

ARTICLE INFO

Article history:
Received 10 March 2022
Revised 31 March 2022
Accepted 4 April 2022
Available online 10 April 2022

Dataset link: Proteomic identification of an endogenous synaptic SUMOylome in the developing rat brain (Original data)
Dataset link: Supplementary files: Proteomics datasets of developing rat brain: synaptic proteome and SUMO2/3-ylome. (Original data)

Keywords:
Brain development
Synapses
Subcellular fractionation
SUMOylome

ABSTRACT

During brain development, synapses undergo structural rearrangements and functional changes mediated by many molecular processes including post-translational modifications by the Small Ubiquitin-like MOdifier (SUMO). To get an overview of the endogenous SUMO-modified proteins in the developing rat brain synapses, our first aim was to characterize the synaptic proteome from rat at 14 postnatal days (PND14), a period that combines intense synaptogenesis, neurotransmission and high levels of SUMO2/3-ylation. In this purpose, we isolated the synaptosomal fraction by differential centrifugation on sucrose percoll gradient and characterized the synaptosomal proteome by nanoLC-MS/MS. Our second aim was to provide a comprehensive list of the SUMO2/3-modified protein in this compartment. We thus performed an enrichment in SUMO2/3-ylated proteins from the synaptosomal fraction by denaturing immunoprecipitation using specific anti-SUMO2/3 antibodies prior to proteomics analysis. The information presented in this article complement the publication "Proteomic Identification of an Endogenous Synaptic SUMOylome in the Developing Rat Brain" [1], by focusing on the characterization of the synaptic proteome of PND14 rat brain. Altogether, these data can inform future experiments focused on studying the functional

* Correspondence: Carole Gwizdek.
E-mail address: gwizdek@ipmc.cnrs.fr (C. Gwizdek).

https://doi.org/10.1016/j.dib.2022.108151

consequences of synaptic SUMOylation regarding synapses structure and function. In addition, they can provide the basis for future mechanistic studies investigating brain pathologies involving altered SUMOylation levels.

Specifications Table

Subject	Neuroscience: Cellular and Molecular
Specific subject area	Synaptic proteomics, post-translational modifications by SUMO
Type of data	Tables Figures Raw data
How the data were acquired	The data were acquired using nanoLC-MS/MS analysis : nanoLC (U3000, ThermoFisher Scientific) using a reversed-phased analytical column 75 µm i.d. x500mm (3 µm, 100 Å) Acclaim PpeMap 100 C18 (ThermoFisher Scientific), coupled to a Q-exactive plus mass spectrometer (ThermoFisher Scientific). The raw data were analyzed using SEQUEST integrated into the Proteome Discovered v2.2 (ThermoFisher Scientific) by searching the data against the Uniprot Rattus norvegicus Reviewed and Unreviewed FASTA database (accessed on 18 July 2018, 35,575 entries).
Data format	Raw Analyzed Filtered
Description of data collection	Synaptosomal fractions were prepared from PND14 rat brains upon differential centrifugation on Percoll-sucrose density gradients. Synaptosomal proteins were separated by SDS-PAGE and subjected to in-gel trypsin digestion prior to LC-MS/MS analysis. Synaptic SUMO2/3-modified protein were purified by specific anti-SUMO2/3 immunoprecipitation on synaptosomal fraction prior to SDS-PAGE separation, in gel trypsin digestion and LC-MS/MS analysis. Concomitant immunoprecipitations using mouse IgG were performed as negative control.
Data source location	The raw data has been collected in the Proteomics/Lipidomics Facility from the Institut de Pharmacologie Moléculaire et Cellulaire, (Valbonne Sophia-Antipolis, France). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE repository with the dataset identifier PXD028804.
Data accessibility	Repository name: ProteomeXchange Data identification number: PXD028804 Direct URL to data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD028804; Mendeley Data, V1, doi: 10.17632/bf99y7dsnh.1
Related research article	Pronot M, Kieffer F, Gay AS, Debayle D, Forquet R, Poupon G, Schorova L, Martin S, Gwizdek C. "Proteomic Identification of an Endogenous Synaptic SUMOylome in the Developing Rat Brain." Front Mol Neurosci. 2021 Nov 23;14:780,535. doi: 10.3389/fnmol.2021.780535. eCollection 2021. DOI: 10.3389/fnmol.2021.780535.

Value of the Data

- The datasets presented here and in the related article [1] provide relevant information regarding the synapse protein content in the developing rat brain and offer an overview of the synaptic SUMO2/3-ylated candidate proteins.
- The present analysis focuses on the characterization of the synaptic proteome of PND14 rat brain, with GO terms enrichment analyses and prediction of synaptic protein/protein clusters highlighted for the SUMO2/3-ylation status of its components.
- Both datasets provide a unique resource to the scientific community to thoroughly assess the SUMO-mediated regulation of synapses structure and function.
- These data could be used/reused to adduce further insights into brain pathologies involving altered SUMOylation levels.

1. Data Description

We recently published two proteomics datasets [1]: one identifies synaptic proteins from PND14 rat brain and the other provides a list of SUMO2/3-ylated protein candidates in the synaptic compartment. In the present article, we provide a further characterization of the synaptic proteome of the PND14 rat brain with enrichment analyses for GO terms using generic (Fig. 2 and Supplementary Table S1) or synapse specific tools (Fig. 3 and Supplementary Table S1) as well as a presentation of the main synaptic protein/protein sub-networks highlighting the SUMO2/3-ylation status of its components (Fig. 4 and Supplementary Table S2) as referenced in the SUMO2/3-ylome we have established. The overall workflow for samples preparation, datasets acquisition and data analysis is depicted in Fig. 1. To map synaptic protein content, PND14 rat forebrains were subjected to differential centrifugation on four-steps Percoll-sucrose density gra-

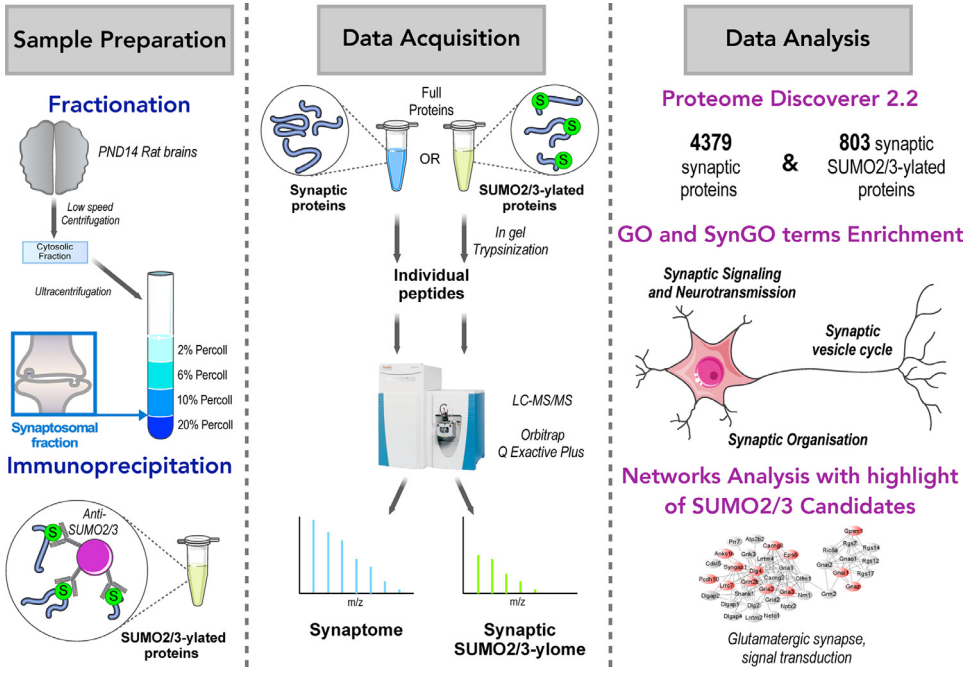

Fig. 1. Experimental workflow for the proteomics analysis of the synaptic proteome and SUMO2/3-ylome.

dients in order to isolate the synaptic compartment. Four independent synaptosomal preparations were analyzed by LC-MS/MS. The MS data were analyzed using Proteome Discoverer v2.2 against the Rattus norvegicus UniProt Knowledgebase. Only "Master Proteins" passing a cut-off of 1% False Discovery Rate for both peptides and proteins (Combined Protein FDR Confidence: "High") were considered for further analysis (Supplementary Table S3). We selected among the filtered proteins those detected in at least 3 of the 4 preparations to establish a PND14 rat brain synaptic proteome thus comprising 4379 proteins. To isolate synaptic SUMO2/3-modified proteins, denaturing immunoprecipitation using a specific anti-SUMO-2/3 antibody combined with a peptidic elution were performed on PND14 rat brain synaptosomal lysates. Four independent synaptic SUMO2/3 immunoprecipitation assays, with concomitant IgG negative controls, were also subjected to LC-MS/MS analysis. For each immunoprecipitation, proteins identified in IgG eluate were withdrawn from the attendant SUMO2/3 identified proteins list. The proteins present at least in 3 of the 4 filtered lists were selected as synaptic SUMO2/3 modified proteins, leading to a final list of 803 candidate SUMO2/3 targets. More than 97% of the synaptic SUMO2/3-ylome overlap with the synaptic proteome [1]. The repository data files located at https://www.ebi.ac.uk/pride/archive/projects/PXD028804 contain: LC-MS/MS raw files generated for these datasets, msf. files generated by Proteome Discoverer analysis and results for identified proteins as xlsx files. The lists of proteins assigned to each dataset are available in the Supplemental Information of the original publication [1]. The PND14 synaptic proteome was subjected to GO terms enrichment analysis using the clusterProfiler tool available on the ProteoRE interface [2] followed by dimensionality reduction using Revigo [3] (Fig. 2 and Supplementary Table S1). The top 20 enriched GO terms for Cellular Components (CC), Molecular Functions (MF) or Biological Processes (BP) are represented in Fig. 2. In line with a synaptic enrichment, terms related to synapses specificities or referring to pathways or organelles tightly connected to synaptic organization and function, such as vesicle-mediated trafficking or mitochondrial processes, are overrepresented. In addition, the identified synaptic proteins were classified based on their association to the synapse specific Gene Ontology terms using SynGO [4]. Sunburst plots of all the significant Synaptic GO terms for CC and BP as well as hierarchical trees representing the top 20 enriched terms in each category are shown in Fig. 3 and detailed in Supplementary Table S1. We observe a high enrichment in pre- and post-synaptic components as well as in many synaptic functions such as vesicle cycle, neurotransmission or synaptic signaling. Last, Fig. 4 illustrates the eight most important protein clusters associated to the proteins we identified in the developing rat brain synapses, along with their associated processes, as predicted by the Cytoscape platform [5] implemented with StringApp [6]. The complete details of predicted protein-protein interactions are provided in Supplementary Table S2. Proteins identified in our synaptic SUMO2/3-ylome are colored in red while other synaptic components appear in gray. We found SUMO2/3-modified proteins in all the clusters, showing the implication of the SUMOylation in all the represented synaptic processes.

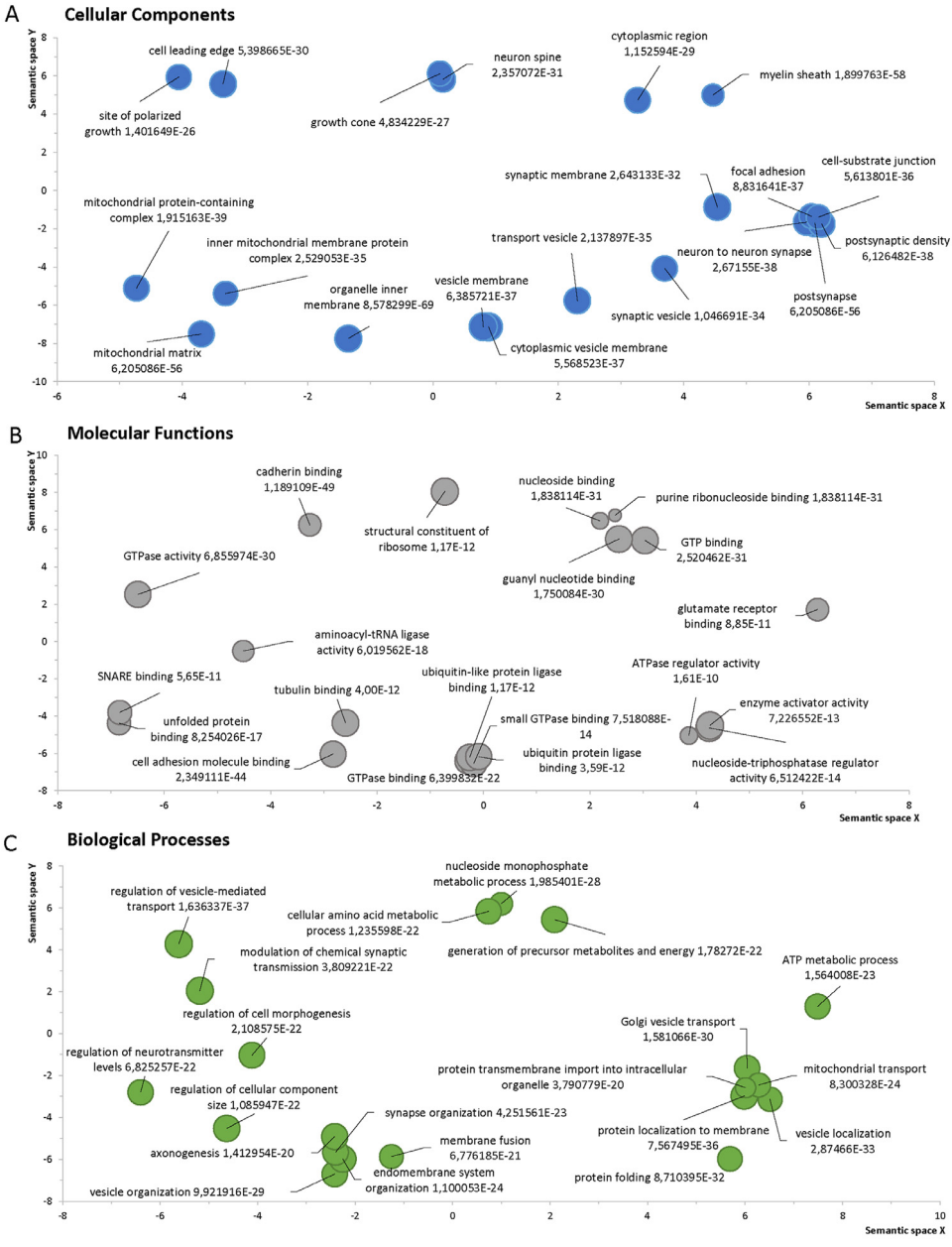

Fig. 2. Terms enrichment analysis of the synaptic proteome according to the Gene Ontology, knowledge base. The list of synaptic proteins was subjected to enrichment analysis against the rat proteome followed by redundancy reduction and top 20 terms are represented in bubble plot for GO Cellular Components (A), GO Molecular functions (B), GO Biological Processes terms (C).

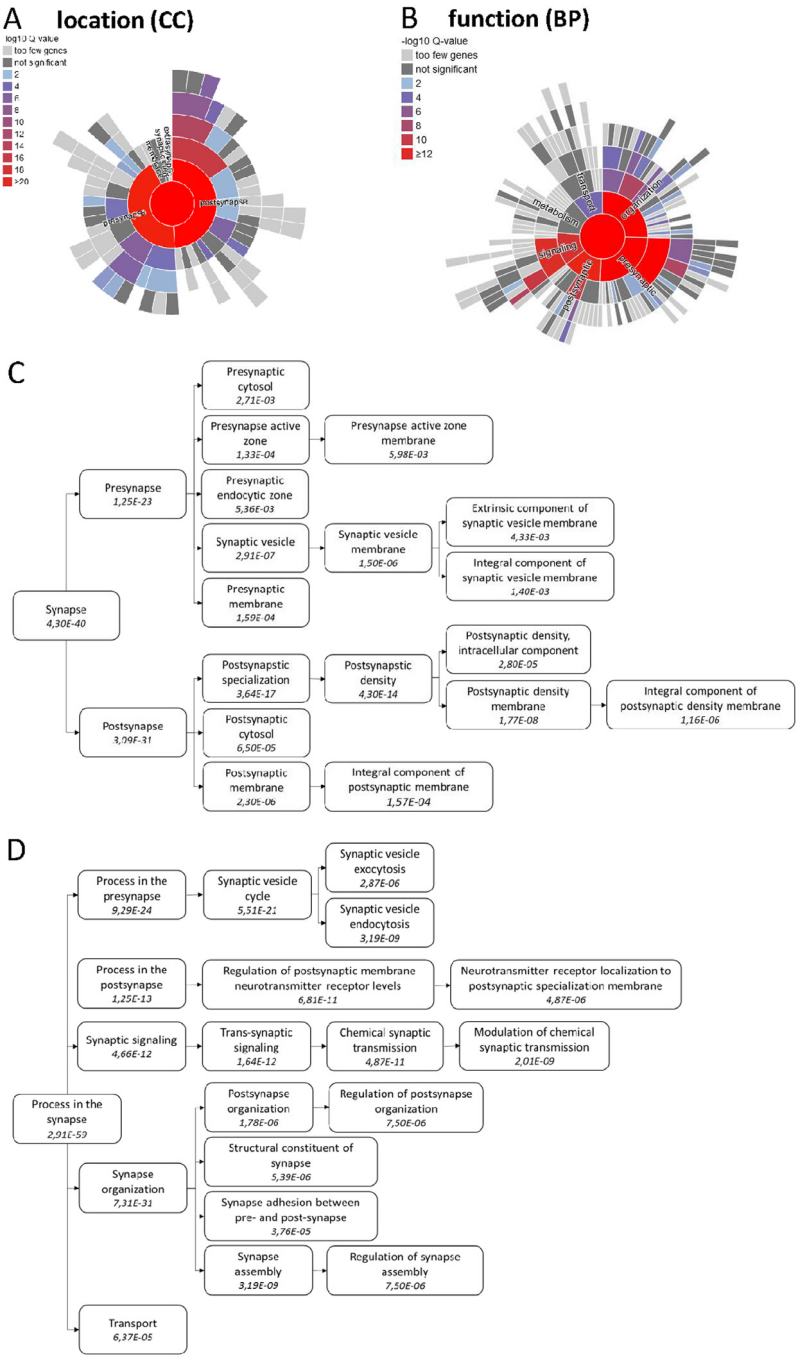

Fig. 3. Terms enrichment analysis of the synaptic proteome according to the synapse specific SynGO database. The list of synaptic proteins was subjected to enrichment analysis the against "brain expressed" background and organized in sunburst plot for all the sugnificant terms (A-B) or in top 20 terms hierarchical tree (C-D) for location/CC (A-C) or function/BP (B-D).

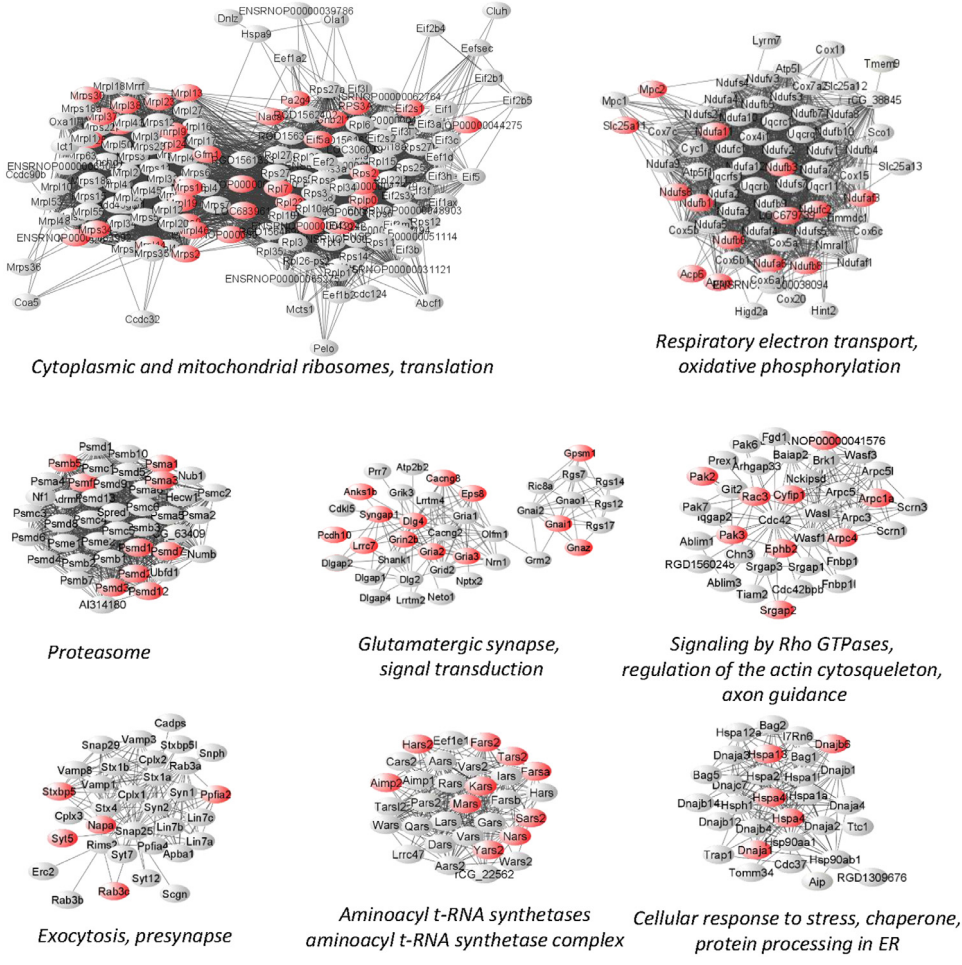

Fig. 4. Subnetwork organization of the identified synaptic proteins: A network was built by the Cytoscape platform using the list of synaptic proteins identified in the present datasets as input and the STRING protein query as Data Source, and further subjected to MCL clustering. The eight first sub-clusters ranked according to their size are presented here. The candidate SUMO2/3 targets are highlighted in red and the other synaptic proteins are coloured in gray.

2. Experimental Design, Materials and Methods

2.1. Experimental design

The experimental workflow we set up, combining synaptosomes isolation, specific SUMO2/3-ylated proteins immunoprecipitation and LC-MS/MS analysis, aims to: (i) map the synaptic proteome from postnatal day 14 rat (PND14) brains and (ii) delineate a list of endogenous SUMO2/3-ylated candidate proteins in this compartment. Each protein dataset arises from four biological replicates. In the present study, the synaptic protein dataset was further characterized by generic or synapse specific GO terms enrichment analysis. Besides, a mapping of predicted protein clusters including the SUMOylation status of their components was generated from the identified synaptic proteins.

2.2. Materials and methods

2.2.1. Biological material and experimental procedures

Synaptosomes from PND14 Wistar rat brain were obtained by differential centrifugation on four-steps sucrose-Percoll gradients as described in the related article [1]. To characterize the PND14 rat brain synaptic proteome, 10 μg of proteins of four independent synaptosomal preparations were separated on gradient SDS-PAGE, in-gel reduced/alkylated by a treatment with DTT/IAA, digested by trypsin and analysed by LC-MS/MS. Regarding the synaptic SUMO2/3-ylome, four independent immunoprecipitations were performed on 8 mg of proteins from denatured synaptosomal lysates using 160 μg of immobilized monoclonal anti-SUMO2/3 antibodies or mouse IgG as negative control. Proteins eluted by peptidic competition were concentrated by TCA precipitation, separated on gradient SDS-PAGE and subjected to in gel reduction/alkylation and trypsin digestion prior to LC-MS/MS analysis.

2.2.2. Data acquisition

To perform sample analysis, the nanoHPLC ultimate 3000 (Thermo Fisher Scientific) was coupled via a nanoelectrospray ionization source to a Hybrid Quadrupole-Orbitrap High Resolution Mass Spectrometer (Thermo Fisher Scientific). 5 µl of peptidic solution was injected and concentrated on a µ-Precolumn Cartridge Acclaim PepMap 100 C18 (i.d. 5 mM, 5 µm, 100 Å, Thermo Fisher Scientific) using a H2O/ACN/FA 98%/2%/0.1% solution at a flow rate of 10 µL/min. Peptides separation was next achieved on a 75 µm i.d. x 500 mM (3 µm, 100 Å) Acclaim PepMap 100 C18 column (Thermo Fisher Scientific) at a flow rate of 200 nL/min with the solvent system: (A) 100% water, 0.1%FA, (B) 100% acetonitrile, 0.08% FA, according to the following gradient: $t = 0$ min 4% B; $t = 3$ min 4%B; $t = 170$ min, 35% B; $t = 172$ min, 90% B; $t = 180$ min 90% B (temperature set at 35 °C). MS spectra were acquired at a resolution of 70 000 (200 *m/z*) using the following settings: scan range of 150−1800 *m/z*, AGC target value of 5e5 and maximum injection time of 50 ms. The 10 most intense precursor ions were selected and isolated with a window of 2 *m/z* and fragmented by Higher energy C-Trap Dissociation with normalized collision energy of 27. MS/MS spectra were acquired in the ion trap using the following settings: resolution of 17 500 (200 *m/z*), AGC target value of 2e5 and maximum injection time of 100 ms.

2.2.3. Data analysis

Raw data were reprocessed using Proteome Discoverer v2.2 (ThermoFicher Scientific) implemented with the SEQUEST HT module and searched against the UniProtKB Rattus norvegicus Reviewed and Unreviewed FASTA database with the following settings: enzyme specificity fixed to trypsin with two missed cleavages allowed, carbamidomethylation of cysteines set as a fixed/static modification and only oxidation of methionine considered as dynamic/variable modification. A mass accuracy of $\pm$10 ppm was used to precursor ions and 0.02 Da for product ions. Results were filtered at 1% FDR for both peptides and proteins as estimated by target-decoy method. To establish the synaptic proteome of PND14 rat brain, only Master Proteins detected in at least 3 of the 4 synaptic preparations were selected. For the synaptic SUMO2/3-ylome, proteins identified for each immunoprecipitation in the mouse IgG eluate were removed in the concurrent SUMO2/3 eluate. Then, proteins present in at least 3 of the 4 filtered lists were selected as synaptic candidates for SUMO2/3 modification.

2.2.4. Bio-informatics analysis

Gene Ontology (GO) terms enrichment analyses for Cellular Components (CC), Molecular Functions (MF) or Biological Processes (BP) were performed against the rat proteome using clusterProfiler (ontology level 2) on the web interface ProteoRE [2] (https://proteore.org/). Hits were selected according to their adjusted p value using the Benjamini- Hochberg method (adjusted p-value $\leq$ 0.01). To reduce redundancy in the GO terms, the clusterProfiler output was fed into Revigo [3] (Reduce + Visualize Gene Ontology, http://revigo.irb.hr/) and p-values were used to select and cluster GO terms with a similarity score of 0.5 (small) for BP and 0.7 (medium) for CC and MF. Enrichment analyses using the synapse specific online data analysis platform SynGO 1.1

[4] (https://syngoportal.org/) were performed against the "brain expressed" background, setting medium stringency. Top-levels terms were used as labels for the sunburst plots and to construct hierarchical trees. Network analysis were performed using the Cytoscape application (v.3.8.2) [5] implemented by the StringApp plug-in [6]. A global confidence score of 0.7 or greater including all edges was used to retrieve interaction data from the STRING database [7]. The resulting network was subjected to MCL clustering at granularity 4.

Ethics Statements

The animal study was complied with the ARRIVE guidelines and were carried out in accordance with EU Directive 2010/63/EU for animal experiments. Protocols were reviewed and approved by the Animal Care and Ethics Committee (Comité Institutionnel d'Ethique Pour l'Animal de Laboratoire N28, Nice, France).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Proteomic identification of an endogenous synaptic SUMOylome in the developing rat brain (Original data) (ProteomeXchange).

Supplementary files: Proteomics datasets of developing rat brain: synaptic proteome and SUMO2/3-ylome (Original data) (Mendeley Data).

CRediT Author Statement

Félicie Kieffer: Investigation, Methodology, Writing – original draft; **Marie Pronot:** Investigation, Writing – original draft; **Anne-Sophie Gay:** Investigation, Writing – original draft; **Delphine Debayle:** Investigation, Writing – original draft; **Carole Gwizdek:** Conceptualization, Methodology, Investigation, Writing – original draft.

Acknowledgments

This work was supported by the 'Fondation pour la Recherche Médicale' for the PhD fellowships #FDT202012010480 to MP and #ECO201906008982 to FK, the 'Fondation pour la Recherche sur le Cerveau' (FRC AO2009) for financial support. This work was also supported by grants from the French government via the 'Investments for the Future' LabEx 'SIGNALIFE' (ANR-11-LABX-0028–01) and the IDEX UCAJedi ANR-15-IDEX-01 as well as the CG06 (AAP santé), the GIS IBiSA and the CSI funding program of Université Côte d'Azur for the funding of the Proteomics/Lipidomics Facility from the Institut de Pharmacologie Moléculaire et Cellulaire.

References

[1] M. Pronot, et al., Proteomic identification of an endogenous synaptic Sumoylome in the developing rat brain, Front Mol. Neurosci. 14 (2021) 780535, doi:10.3389/fnmol.2021.780535.

[2] L. Nguyen, et al., Designing an in silico strategy to select tissue-leakage biomarkers using the galaxy framework, Methods Mol. Biol. 1959 (2019) 275–289, doi:10.1007/978-1-4939-9164-8_18.

[3] F. Supek, et al., REVIGO summarizes and visualizes long lists of gene ontology terms, PLoS ONE 6 (7) (2011) e21800, doi:10.1371/journal.pone.0021800.
[4] F. Koopmans, et al., SynGO: an evidence-based, expert-curated knowledge base for the synapse, Neuron 103 (2) (2019) 217–234 e4, doi:10.1016/j.neuron.2019.05.002.
[5] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, doi:10.1101/gr.1239303.
[6] N.T. Doncheva, et al., Cytoscape stringapp: network analysis and visualization of proteomics data, J. Proteome Res. 18 (2) (2019) 623–632, doi:10.1021/acs.jproteome.8b00702.
[7] D. Szklarczyk, et al., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic. Acids. Res. 45 (D1) (2017) D362–D368, doi:10.1093/nar/gkw937.