Analyzing the regulation of synaptic vesicle mobility by pre-synaptic liquid phase-separation in health and disease

A Data Management Plan created using DMPonline.be

Creator: Gokhan Özturan

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1165923N

ID: 199955

Start date: 07-03-2022

End date: 26-06-2027

Project abstract:

Loss of synaptic function closely associates with cognitive impairment. Work from my host lab shows that in dementia, the protein Tau abnormally invades synapses, where it binds to synaptic vesicles. However, it is not known how Tau causes synaptic degeneration. My hypothesis is that the invasion of Tau into synapses disrupts the healthy synaptic vesicle pool, causing decline. Previous work indicates that under normal conditions, synapses pack surplus vesicles in ‘phase-separated droplets’, however the extra interactions provided by pathological Tau that my host lab uncovered, could disrupt this balance. I will determine the role of Tau, and its binding partners in liquid-liquid phase separation of synaptic vesicles using advanced biophysical methodology including confocal imaging and optical tweezer measurements and define key protein interaction surfaces that are required to drive this process. I will then use clever genetic manipulations to interfere with the process of phase separation of vesicles in vivo in Drosophila. I will ask if this counteracts Tau-induced synaptic dysfunction and synaptic decline based on electrophysiology and electron microscopy imaging of synaptic terminals. If successful, my project will have defined a new role for Tau directly in liquid liquid phase separation of the synaptic vesicle pool and I will know if interfering with phase separation of synaptic vesicles counteracts Tau-induced synaptic dysfunction and degeneration.

Last modified: 16-06-2023

Analyzing the regulation of synaptic vesicle mobility by pre-synaptic liquid phase-separation in health and disease
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, ... • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Data Set 1.1	Confocal Images, Gel Images	Generate New Data	Digital	Experimental	tif	<50 TB	20 TB
Data Set 1.2	Recordings of Live Imaging	Generate New Data	Digital	Experimental	tif	<50TB	20 TB
Data Set 1.3	Electrophysiology recordings	Generate New Data	Digital	Experimental	txt	<1 GB	1 GB
Data Set 1.4	Drosophila melanogaster Lines	Generate New Data	Physical	Experimental	NA	NA	NA
Data set 1.5	Protein expression plasmids	Generate New Data	Physical	Experimental	NA	NA	NA
Data Set 1.6	Purified Synaptic Vesicles	Generate New Data	Physical	Experimental	NA	NA	NA
Data Set 1.7	Previous Confocal Images	Reuse the existing data	Digital	Experimental	tif	<50 TB	1-2 TB
Data set 1.8	Pre-existing drosophila lines	Reuse existing data	Physical	Experimental	NA	NA	NA

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

The data from the previous research in the lab is going to be used. The details are given below:
- Zhou, L. *et al.* (2017) Tau association with synaptic vesicles causes presynaptic dysfunction. *Nat. Commun.* **8,** 15295 DOI: 10.1038/ncomms15295
- McInnes, J., et al., (2018). Synaptogyrin-3 Mediates Presynaptic Dysfunction Induced by Tau. *Neuron*, *97*(4), 823–835.e8. DOI: 10.1016/j.neuron.2018.01.022
-Largo-Barrientos, P., et al. (2021). Lowering Synaptogyrin-3 expression rescues Tau-induced memory defects and synaptic loss in the presence of microglial activation. *Neuron*, *109*(5), 767–777.e5. DOI: 10.1016/j.neuron.2020.12.016

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, animal data

I will collect brain from wild type and transgenic mouse for isolation of synaptic vesicles. The ethical committee approval is already established for the project (ECD project number 194/2018)

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven and VIB in accordance with the framework agreement of both institutes. VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application it will be planned so that publications need not be delayed. The use of plasmids, proteins and other materials will be subjected to the terms described in their respective MTAs.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook that refer to specific datasets. All datasets will be accompanied by a README.txt file containing all the associated metadata.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The metadata will be based on a generalized metadata scheme with the following elements:
- Title: The title of the experiment
- Creator: Last name, first name, organization
- Date and time reference
- Subject: Choice of keywords and classifications
- Description: Explanation related to the content of the data set and other information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
Format: Details of the file format
Resource Type: data set, image, audio, etc.
Identifier: DOI (when applicable)
Access rights: closed access, embargoed access, restricted access, open access.
For specific datasets, additional metadata will be associated with the data file as appropriate (e.g., evaluated regions on images as zipped ROI files).
The final dataset will be accompanied by this information under the form of a README.txt document. This file will be located in the top level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory.

3. Data storage & back-up during the research project

Where will the data be stored?

Digital files will be stored on KU Leuven servers, except for private data that will be stored on KU Leuven secure server (digital vault).
- Tissue samples: Tissues will be stored locally in the laboratory.
- Vectors: Plasmids will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request.
- Drosophila lines will be stored in a dedicated room and managed using a specific database for storage of the corresponding information (including genotype, origin, number of vials and date of transfer, crossing schemes) and vial tracking via unique QR codes.
- Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.
- Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in the KU Leuven servers.
- Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers. Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database.

How will the data be backed up?

KU Leuven drives are backed-up according to the following scheme:
- data stored on the “L-drive” is backed up daily using snapshot technology, where all incremental changes in respect of the previous version are kept online; the last 14 backups are kept.
- data stored on the “J-drive” is backed up hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept.
- data stored on the digital vault is backed up using snapshot technology, where all incremental changes in respect of the previous version are kept online. As standard, 10% of the requested storage is reserved for backups using the following backup regime: an hourly backup (at 8 a.m., 12 p.m., 4 p.m. and 8 p.m.), the last 6 of which are kept; a daily backup (every day) at midnight, the last 6 of which are kept; and a weekly backup (every week) at midnight between Saturday and Sunday, the last 2 of which are kept.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

There is sufficient storage and back-up capacity on all KU Leuven servers:
- the “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.
- the “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Both the “L-drive” and “J-drive” servers are accessible only by laboratory members, and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.
Access to the digital vault is possible only through using a KU Leuven user-id and password, and user rights only grant access to the data in their own vault. Sensitive data transfer will be performed according to the best practices for “Copying data to the secure environment” defined by KU Leuven. The operating system of the vault is maintained on a monthly basis, including the application of upgrades and security patches. The server in the vault is managed by ICTS, and only ICTS personnel (bound by the ICT code of conduct for staff) have administrator/root rights. A security service monitors the technical installations continuously, even outside working hours.
All private data will be rendered anonymous before processing outside the digital vault. Only the PI will be granted access to the server to deposit private data. The PI will be the only responsible for linking patient information, survey data and/or tissue samples, and will strictly respect confidentiality. All de-identified data will be exported from the database by the PI, and stored on KU Leuven servers from where it can be accessed by the research and technical staff from the laboratory.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The total estimated cost of data storage during the project is 35 000. This estimation is based on the following costs:
Digital data (~30TB/year) is stored on "L-drive" : 173,78€/TB/Year for the “L-drive”
Maintaining a Drosophila line alive costs about 4 euro per year including the fly food, vials and personnel’s cost. For this project, there are around 1000 stock vials. The PI and the lab manager if applicable are responsible for the preservation of fly lines as well as administrative and experimental data. All published lines will be preserved for the remainder of the PI’s research career. All unpublished lines will be preserved for a minimum of 5 years after the end of the project.
-Maintaining a mouse colony alive costs about 1,200 euro per year (for 6 cages), excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R’s rule (https://www.nc3rs.org.uk), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimen are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro.
Electricity costs for the -80° freezers present in the labs are included in general lab costs. Data storage and backup costs are included in general lab costs.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

The minimum preservation term of 5 years after the end of the project will be applied to all datasets.

Where will these data be archived (stored and curated for the long-term)?

As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication.
For all other datasets, long term storage will be ensured as follows:
- Digital datasets: files will be stored on the “L-drive”.
- Tissue samples: Tissues will be stored locally in the laboratory.
- Omics data: datasets will be stored on the “L-drive” or, for larger datasets, on the Vlaams Supercomputer Centrum.
- Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C).
- Genetically modified organisms: Drosophila lines will be housed locally.
- Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The total estimated cost of data storage during the 5 years after the end of the is 35 000. This estimation is based on the different costs described previously.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Other, please specify:

All the data sets will be open-access after publishing in peer-reviewed journals, including supplemental information.

If access is restricted, please specify who will be able to access the data and under what conditions.

The plasmids, drosophila lines and other data will be available upon request by email.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific research outputs are detailed below:
- Vectors: Upon publication, all vectors supporting a manuscript will be made publicly available via the non-profit plasmid repository Addgene, along with the corresponding DNA sequences. Addgene in turns performs quality control on the DNA, curates the plasmids online with all relevant information (maps, sequences), and for a minimal cost (typically $65) ships the vectors upon simple request and signature of a material transfer agreement. The MTA will be prepared before depositing the vectors with the help of our organization's Tech Transfer office. For transfer between nonprofit or academic institutions, Addgene typically uses the Uniform Biological Material Transfer Agreement . All non-published vectors and the associated documentation will be shared by the PI upon request and after signature of a material transfer agreement, at no cost except the cost of shipment.
- Genetically modified organisms: All genetically modified organisms used in publications will be made available to researchers upon request at the time of publication.
- Antibodies, synthetic and recombinant compounds: samples will be stored as appropriate in the laboratory. Within availability, they will be shared with interested researchers upon request.
- Manuscripts: All scientific publications will be shared openly. Manuscripts submitted for publication will be deposited in a pre-print server such as bioRxiv. At the time of publication, research results will be summarized on the PI's website (https://verstrekenlab.sites.vib.be/en) and post-print pdf versions of publications will be made available there if allowed by copyright agreements. Publications will also be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.
- Nucleic acid and protein sequences: Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences).
- Data that do not support publication will be either deposited in an open access repository or made available upon request by email.

When will the data be made available?

The data will be made available upon publication of the research results
As a general rule all research outputs will be made openly accessible at the latest at the time of publication.

Which data usage licenses are you going to provide? If none, please explain why.

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a nondisclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

Expected costs for data sharing are not determined yet. The costs for data sharing will be paid by the researchers who request the data from the lab. Other publicly available data at Addgene or Bloomington Drosophila Center charge for sharing the data.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Gökhan Özturan

Who will manage data storage and backup during the research project?

Patrik Verstreken

Who will manage data preservation and sharing?

Patrik Verstreken

Who will update and implement this DMP?

Gökhan Özturan

Analyzing the regulation of synaptic vesicle mobility by pre-synaptic liquid phase-separation in health and disease
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

In the project previously generated data will be reused and also new data will be generated.
The Research will reuse previously generated animals: wild type and synaptogyrin-3 knock out mouse, Drosophila (variety of genetically modified lines, bacteria lines and antibodies.
Biological Samples : The Tau expression plasmids with different variants of MAPT gene for protein expression, The plasmid with codon optimized Synapsin gene, new genetically modified animals (Drosophila with phase separation specific mutations generated for the project), DNA sequences from these previously mentioned plasmids and animals, frozen tissue samples in cryovials, samples prepared from live animals stored at 4°C, -20°C, and -80°C.
Experimental data: Digital images from the Confocal Microscopy Images (FRAP experiments, live imaging) both for in vivo and in vitro experiments, Phase separation assays (droplet formation, dispersion, mobility assays); gel scans and western blot images related to protein purification and expression.
Estimated Size of the data
DNA and protein sequences : format = text-based format (.fasta/.fa)
Confocal imaging: Format= Tif Files, PNG, JPEG Size = 10mb-200 mb/per image, in total 20 TB
Confocal Live Imaging : Format = tif files, Size = 1GB -2 GB/per recording, in total 20 TB

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person : Prof. Patrik Verstreken
2. Storage capacity/repository
    - during the research

      Digital files will be stored on KU Leuven servers, except for private data that will be stored on KU Leuven secure server (digital vault).There is sufficient storage and back-up capacity on all KU Leuven servers.

      KU Leuven drives are backed-up according to the following scheme:
      - data stored on the “L-drive” is backed up daily.
      - data stored on the “J-drive” is backed up hourly and weekly.

      The following types of data will be stored as follows:
      - Tissue samples: Tissues will be stored locally in the laboratory.
      - Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request.
      - Bacterial strains will be stored in a -80°C freezer in the lab of Patrik Verstrken. Costs are covered by general lab expenses.
      - Genetically modified organisms (mouse): Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. All animals will be registered in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, ethical approval documents and animal provider receipts.
      - Genetically modified organisms (Drosophila): pre-existing and newly generated lines will be stored in a dedicated room and managed using a specific database for storage of the corresponding information (including genotype, origin, number of vials and date of transfer, crossing schemes) and vial tracking via unique QR codes.
      -Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.
      - Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).
      - Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers. Upon publication, all sequences supporting a manuscript will be made publicly available.

    - after the research

      The minimum preservation term of 5 years after the end of the project will be applied to all datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for “Large volume-storage”) will be covered by Verstreken Lab.

      As a general rule, datasets will be made openly accessible whenever possible at the latest at the time of publication.

      For all other datasets, long term storage will be ensured as follows:
      -Digital datasets: files will be stored on the “L-drive”.
      -Tissue samples: Tissues will be stored locally in the laboratory.
      -Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C).
      -Genetically modified organisms: Drosophila lines will be housed locally. All other lines that are no actively used for experiments will be cryopreserved.
      -Genetically modified organisms (mouse): Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven.
      -Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

I do not plan on deviating from the principle of preservation of data and of the minimum preservation term of 5 years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

No.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Nothing else at the moment.