The Role of Parkinsonism genes in synaptic autophagy

A Data Management Plan created using DMPonline.be

Creators: Patrick Vandormael, WIM VERSEES, Patrik Verstreken https://orcid.org/0000-0002-5073-5393

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: WIM VERSEES, Patrik Verstreken https://orcid.org/0000-0002-5073-5393

Project Administrator: Patrick Vandormael

Grant number / URL: G031324N

ID: 206061

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

Parkinsonism is without a cure and affects millions worldwide. The disease is clinically defined by the occurrence of typical motor symptoms and is commonly associated with the aggregation of the protein alpha-synuclein. While idiopathic cases are often associated with numerous risk factors, familial cases are caused by mutations in >22 different genes. Many of these genes point to a variety of underlying molecular causes and our recent research has identified a cluster of genes that code for proteins involved in protein homeostasis, including LRRK2, SYNJ1, EndoA1, Rab39, and DNAJC6. It is however not clear how these proteins act in a network, how pathogenic mutations affect proteostasis and how to counteract the resulting dysfunction. This project will combine our expertise in protein biochemistry and structure determination with in vivo genetics and structure-function cell biology to mechanistically define how these proteins converge in a functional network, and to develop strategies to reverse pathogenic effects. We will create a unique new collection of nanobodies to facilitate this research and use new screenable fly models to reveal how mutations in these Parkinsonism proteostasis genes cause neuronal cellular dysfunction and proteostasis defects.

Last modified: 01-07-2024

The Role of Parkinsonism genes in synaptic autophagy
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:* •	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Task 1.1, 2.1, & 3.3: whole-brain histology	detection vacuole formation due to neuronal death	new	digital	Experimental measurement	- raw data: Carl Zeiss Image files (.czi), Nikon Image file (.nd2) - Digital images in raster formats: uncompressed TIFF (.tif/.tiff) -metadata: textual data (.rtf, .xml, .txt)	<100MB	
Task 1.1, 2.1, & 3.3: electroretinograms (ERGs)	electrophysiological assay to assess synaptic communication in the adult fly brain	new	digital	Experimental measurement	- Applied Biosystems Sequence Tracer Sequence Trace (.abf)		
Task 1.1, 2.1, & 3.3: autophagosome dynamics	live imaging of mutant larval NMJs	new	digital	Experimental measurement	- raw data: Carl Zeiss Image files (.czi, .lsm), Nikon Image file (.nd2) - Digital images in raster formats: uncompressed TIFF (.tif/.tiff) - Digital video data: MPEG-4 High Profile (.mp4); Audio Video Interleave (.avi) -metadata: textual data (.rtf, .xml, .txt)	<1TB	

Tasks 1.1, 2.1, 2.2, & 3.3: (super-resolution) fluorescence microscopy	including immunofluorescence stainings, overall proteostasis,	new	digital	Experimental measurement	- raw data: Carl Zeiss Image files (.czi, .lsm), Nikon Image file (.nd2) - Digital images in raster formats: uncompressed TIFF (.tif/.tiff) - Digital video data: MPEG-4 High Profile (.mp4); Audio Video Interleave (.avi) -metadata: textual data (.rtf, .xml, .txt)	<1TB	
Task 1.1, 2.1, & 3.3: fly lines	KO/KI collection, crossed with fluorescent sensors, human alpha-syn	existing and new	physical	NA	NA	NA	2 vials per line
Task 1.2: expression plasmids PD proteins	E.coli and mammalian expression plasmids for Parkinsonism proteins	existing and new	physical	NA	NA	NA	Biological and chemical samples: samples stored at -20°C (DNA) and -80°C (cell stocks)
Task 1.2: purified PD proteins	Samples of purified proteins in buffer	new	physical	NA	NA	NA	Samples of purified proteins in buffer are stored at -80°C
Task 1.2, 2.3 and 3.2: data resulting from biochemical and biophysical experiments	Data resulting from SEC-MALS, mass photometry, CD, DSF, MST, BLI and ITC experiments and enzymatic assays	new	digital	Experimental measurement	exported and saved as .csv or .txt files	< 10TB	
Task 1.3: Nb array	Nb phage display library and selected Nb candidates	new	physical	NA	NA	NA	Libraries are stored in dedicated repositories at -80°C
Task 2.1: CLEM	correlative light and electron microscopy	new	digital	Experimental measurement		<1TB	

Task 3.1: atomic-resolution protein structures	structures of single proteins or complexes using X-ray crystallography and/or cryogenic-EM	new	digital	experimental measurement	- EM images are saved as .jpeg or .tif - Format of X-ray diffraction data are dependent on the specific detector used in the synchrotron - Processed data (.mtz; .mrc; .hdf; .pdb; mmCIF) are saved in file formats that are standardly used in the field and required by the commonly accepted data repositories for structural biology data (e.g. the protein Data Bank - http://www.ebi.ac.uk/pdbe/).	> 50TB	

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:

This study will (re)use datasets and tools previously generated by the Verstreken and Versées labs, such as the Parkinsonism KO fly collection, the PD mutant KI fly collection, several fluorescent reporter flies, and Nbs against lead candidates.

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

This study uses in vitro biochemical/biophysical methods, and Drosophila as a model organism which poses no ethical issues.

The experiments involving immunization of a **llama** with the aim of eliciting an immune response and generating Nanobodies, will be performed in collaboration with the group of Jan Steyaert, and are being covered by a general ethical approval concerning the generation of an antigen-specific immune response in Camelidae that is already in place.

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, VUB and VIB in accordance with the framework agreement between the involved universities (KU Leuven or VUB) and VIB. 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.

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

No third-party agreement restricts dissemination or exploitation of the data or strains generated from this project. Existing agreements between VIB, KU Leuven and VUB do not restrict publication of data.

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).

The organisation of research data follows a hierarchical **folder structure** based on model organism (eg. flies/cells/...). At the second level, these folders contain subfolders per experiment type (eg. histology, ERG, CLEM, immunofluorescence,...). At the third level, these folders contain subfolders per individual experiment (with date indication, see below). Each of these experiment folders contain subfolders for raw data, processed data, and analysed data.

Digital files and folders of individual experiments will be named following a standard **naming convention**. All file names in a given dataset will follow the same format: they will start with the date, followed by the project acronym, a concise descriptive name, and an optional version number. Whenever possible, names will be kept under 32 characters, containing only letters, numbers, and underscores. Dots will only be used for version control indicators (minor revisions indicated by decimal numbers, and major revisions by whole numbers): YYYYMMDD_Project_Experiment_ version.format. Similarly, folder naming for individual experiments follows the following structure: YYYYMMDD(_Project)_Experiment

Data will be accompanied by **documentation** containing all contextual and descriptive features of the research data, which allow to understand and (re)use the data. Detailed procedures of each experiment are logged in the electronic laboratory notebook (E-notebook), with reference of each experiment to the directory path of the folder containing raw and analysed data files. This also includes data collection methods, protocols, and code explanation. Documentation is stored at the study- and the data-level, providing data provenance from the original source data to specific datasets linked to publications. Data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database and E-notebook.

Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the E-notebook and/or in hard copy lab notebooks that refer to specific datasets.

To allow long term access and use of research data will be stored or converted to **open file formats** as much as possible.

• Containers: TAR, ZIP
• databases: XML(xlsx), CSV, JSON
• Statistics: DTA, POR, SAS, SAV
• Images: TIFF, JPEG 2000, PNG, GIF
• Tabular data: CSV, TXT
• Text: XML(docx), PDF/A, HTML, JSON, TXT, RTF
• Sequencing data: FASTA, FASTQ

We use **controlled vocabularies** or ontologies when applicable to provide unambiguous meaning, for example:

• Gene Onotology: molecular function, cellular component, and biological role of RNA seq
• ENSEMBL or NBCI identifiers: gene identity
• HUGO Gene Nomenclature Committee: names and symbol of human genes
• FlyBase: names and symbol of Drosophila genes

• UniProt protein accessions: protein identity

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

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 (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets. All datasets will be accompanied by a README.txt file containing all the associated metadata, which will include the following elements:
• Title: free text
• Creator: Last name, first name, organization
• Date and time reference
• Subject: Choice of keywords and classifications
• Structure: internal structure of the dataset, or the meaning of abbreviations (not necessary when it is clear from the in-file documentation).
• Description: Text explaining the content of the data set and other contextual 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.
Additionally, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards more specific to our data type. For specific datasets, additional metadata will be associated with the data file as appropriate.

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

Where will the data be stored?

Digital data
- KU Leuven

- All scientists use an **ELN** to log details of their experiments (history of experiments, protocols and analysed data). Files containing raw and analysed data are stored on KU Leuven servers.
- Primary storage for **active digital files** will be on KU Leuven servers. KU Leuven offers fast ("J-drive) and slower ("L-drive") storage that allows reading/writing/modification of non-confidential, confidential, and strictly confidential data.
- KU Leuven further offers the ManGO platform for storage and management of **large volumes of active research data**. This platform allows secure storage, manual and automated metadata coupling, data workflows, and file sharing.
- Data that is no longer active, can be **archived** on the KU Leuven K-drive, which allows reading of non-confidential, confidential, and strictly confidential data

- VUB

- All scientists use an **ELN**. All raw data of biochemical and biophysical measurements, history of experiments, protocols and analysed data are uploaded to this E-Notebook (for very large files > 10Mb a cross-reference is made to the relevant Sharepoint and hard disk folder). These are backed up in the cloud.
- All data are additional be stored on **VUB Sharepoint**.
- Due to the large amount of data, all raw data of successful structural biology experiments (EM, X-ray) are saved on **dedicated hard disks**. Two copies are made on separate hard disks that are kept on physically different locations. Hard disks are numbered and a list (excel file of data stored on each disk is kept at a central point on the Sharepoint folder). Moreover, a cross-reference is made between the metadata in the ELN and the location of the raw data on the hard disks.
- Additionally, all raw structural biology data are stored at the synchrotron or at the VIB-VUB facility for Bio Electron Cryogenic Microscopy (BECM).

Physical samples

- 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 bacterial 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.
- Fly line stocks are preserved as a minimum of two separate cultures, each maintained at 18°C on a 4-to-5-week generation cycle.

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 in ManGO: Snapshots are made at regular intervals (hourly, daily and monthly) in case data needs to be recovered. The data itself is synchronized on two separate hardware storage systems, each 6 PB large, located at Leuven and at Heverlee (ICTS). The data is protected against calamities at either site by synchronizing it in real-time at hardware level.

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

KU Leuven servers offer sufficient storage for active data (J/L-drive, ManGO) and archived data (K-drive). Required data-storage volumes can be easily scaled up.

Sufficient space on ELN and on **VUB** Sharepoint (5 TB) is available for all processed data. Raw structural biology data is saved on hard disks. New hard disks will be purchased if space is running out. Additionally, all raw structural biology data are stored at the synchrotron or at the VIB-VUB facility for Bio Electron Cryogenic Microscopy (BECM) where enough storage space is available

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

The **buildings** or access to the labs on the KU Leuven and VUB campus are restricted by badge system so only employees are allowed in and visitors are allowed under supervision after registration.
Access to **KU Leuven servers** is possible only through using a KU Leuven user-id and password, and user rights only grant access to their own data, or data that was shared to them. Data in these drives 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.
VUB Sharepoint is system encrypted and uses user-authentication via VUB email address and password. Read and write permission of the SharePoint site will be closely monitored so that only authorized people will have access to dedicated folders on the SharePoint site. Hard disks used for back-up will be stored in locked cabinet and system encryption will be put in place (e.g. using BitLocker).

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

-The costs of digital data storage at **KU Leuven** are as follows:

- 519€/TB/Year for the “J-drive” (fast storage, active data),
- 569,2€/5TB/Year for the “L-drive” (medium speed storage, active data),
- 35€/TB/Year for the ManGO platform (large volume storage, active data),

VUB SharePoint is freely available. Costs of hard disks (about 130 Euro / 4 TB): total costs < 500 Euro for entire project.
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...).

According to KU Leuven and VUB RDM policy, relevant research data will be preserved on the university's servers for a minimum of 10 years. Such data include data that are at the basis of a publication, that can only be generated or collected once, that are generated as a result of a substantial financial or personal effort, or are likely to be reused within the research unit or in wider contexts.

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

As a general rule all research outputs (data, documentation, and metadata) related to publications will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org). We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a separate data repository.
Other research data will be archived on KU Leuven and VUB servers as described above.

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

-The costs of digital data storage on KU Leuven servers are as follows:

- 113,84€/TB/Year for the “K-drive” (archive, inactive data).
- 569,2€/5TB/Year for the “L-drive” (medium speed storage, active data)
- 35€/TB/Year for the ManGO platform (large volume storage, active data)

VUB SharePoint is freely available. Costs of hard disks (about 130 Euro / 4 TB): total costs < 500 Euro for entire project.
Data storage and backup costs are included in general lab costs.

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

- Source data for **electrophysiology** experiments, and **binding and kinetics** experiments will be deposited on the online repository Zenodo, where they will be assigned a unique and persistent identifier.
- Experimentally determined **protein structures** from X-ray crystallography and cryo-EM will be deposited on the Electron Microscopy Data Bank (EMDB) and RCSB Protein Data Bank (RCSB PDB), where they will be assigned a unique and persistent identifier.
- To ensure data findability, links and references these datasets, workflows and modes will be included in the **data availability statements** of the associated publication.

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

N/A

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.

- Yes, Intellectual Property Rights

- The researchers involved and the IP team of the VIB TechTransfer Office shall make the necessary arrangements in order to maintain an embargo on the public access of research data, at least until the essential steps in securing intellectual property (e.g. the filing of a patent

application) have been taken. 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.

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

- Source data for electrophysiology experiments, and biochemical and biophysical experiments will be deposited on the online repository Zenodo, where they will be assigned a unique and persistent identifier.
- Experimentally determined protein structures from X-ray crystallography and cryo-EM will be deposited on the Electron Microscopy Data Bank (EMDB) and RCSB Protein Data Bank (RCSB PDB), where they will be assigned a unique and persistent identifier.

When will the data be made available?

All research outputs (data, documentation, code, and associated metadata) will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

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

Data is typically available 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. Software and code usually are available under a GNU General Public License or an Academic Non-commercial Software License.

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?

It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be covered by the laboratory budget.

6. Responsibilities

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

The researchers who generate the data are responsible for managing data, documentation, and metadata.

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

The researchers who generate the data are responsible for storage and backup, with support from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.

Who will manage data preservation and sharing?

The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René

Custers and Alexander Botzki for the electronic laboratory notebook (ELN), and from Raf De Coster for the KU Leuven drives.

Who will update and implement this DMP?

The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP, with support from Patrick Vandormael.