FWO DMP 1SE3622N

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

1. General Information	
Name applicant	**Ruggiero Pio Cassatella**
FWO Project Number & Title	**1SE3622N - Discovery of endogenous and recombinant stabilizers of Gamma-Secretase-APP complexes to provide mechanistic underpinnings for Alzheimer's disease therapy**
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☒ Reuse existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	Human genetics has shown that mutations in three genes- *PSEN1, PSEN2 and APP*- are causative of early-onset familial Alzheimer's disease (FAD). Presenilin (PSENs) constitutes the catalytic subunit of an intramembrane protease complex, named gamma-secretase (GSEC), and the amyloid precursor protein (APP) represents one of its substrates. GSECs cut sequentially APP generating amyloid-β (Aβ) peptides of different length. Numerous studies have shown that pathogenic mutations enhance the generation of longer, partially digested and toxic Aβ peptides. My host lab has shown that the pathogenic effects of AD-causing mutations stem from the destabilization of GSEC-APP (Enzyme-Substrate, E-S) interactions during the sequential proteolysis. Notably, a remarkable linear correlation between the magnitude of the mutation-induced destabilization and the age of onset strongly supports the pathophysiological relevance of these findings. Furthermore, these studies show that changes in the molecular environment surrounding GSEC-APP complexes can modulate their stability. We propose that endogenous modulators of GSEC affecting E-S interactions may promote or reduce the production of longer and toxic Aβ peptides, therefore affecting an individual's risk of developing AD. Thus, I will investigate GSEC-interactome in health and disease conditions, and tackle GSEC metastability using Nanobody (Nb)-based technology with three major aims. (i) Conformation-specific Nbs binding to GSEC and/or GSEC-APP complexes will allow the generation of Nbs binding with high affinity to the GSEC complexes. (ii) Discovered Nbs will be fused with TurboID biotin ligase to uncover endogenous interactors of GSEC in human brain in disease and health conditions, using proteomics. The discovery of newly endogenous modulators may point out novel pathways contributing to disease pathogenies. The generated insight may offer alternative therapeutic targets for AD treatment. At the same time, the generation of Nb-chaperons with potentially stabilizing properties may provide prototype molecules for translation research. **Digital images:** The hereby indicated dataset includes digital images acquired at fluorescent microscopy on the different cell models used; digital images of western blots, gel scans; pictures and illustrations obtained from experimental data sets. Data formats: -Digital images: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), Adobe Portable Document Format (.pdf), bitmap (.bmp)

	- Quantitative tabular data: MS Excel (.xls/.xlsx), Graph pad, tab-delimited file (.tab) **Vectors:** Bacterial vectors, mammalian expression vectors, viral vectors will be used to express nanobodies, GSEC and APP. Several DNA libraries of newly identified Nanobodies will generated during this project to identity hits with high affinity vs stabilizers. Data formats: -Nucleotide sequences: raw sequence data trace (.ab1), text-based format (.fasta) -Text files describing vectors and inserts: plain text data (.txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) **Omics data:** This study involves the generation of cDNA libraries from Nanobody pannings/screenings (aim 1) as well as proteomics data derived from the analysis of the endogenous interactome of GSEC in brain (aim 2). Accordingly, the following data (and formats) will be generated during this project: Data formats: - DNA, protein and peptide sequence: raw sequence data trace (.ab1), text-based format (.fasta/.fa) - Text files describing omics analysis: plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) - List of peptides are represented as quantitative tabular data: comma-separated value files (.csv), MS Excel (.xls/.xlsx) - Text files describing models and storage information: plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) **Cell lines:** Bacterial strains for the production of expression vectors (DNA) and Nanobody expression (Protein), immortalized human neuronal progenitors (ReNcell VM), immortalized human cell line (HEK293) required

	for Aim1 and Aim2 of the proposed study, insect cells (Hi5 cells) required for GSEC and APP production and purification. Cryotubes of biological samples stored at -80˚C will be labelled with reference number that links to an entry in strain database. The stable cell lines will be stored at -180˚C. Research documentation generated from online sources and from collaborators, including publications, tutorials, laboratory notes and protocols will be collected in text files such as plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf) Data formats: - Biological samples: frozen cell lines (-80˚C, liquid nitrogen), frozen cell pellets and cell-membranes (-20˚C), bacterial glycerol stocks, viral particles. - Text files describing the different cell lines and samples derived thereupon: plain text data (Unicode, .txt), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf).

3. Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☐ Yes ☒ No If yes: - Privacy Registry Reference: - Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s).	☒ Yes ☐ No If yes: - Reference to ethical committee approval: I will generate cell membranes from post-mortem human brain samples to discover endogenous interactors of gamma-secretase via Nanobody-based – TurboID "in membrane" labelling. Ethical committee number approval: E-2021-2759
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	☒ Yes ☐ No If yes, please comment: We do not rule out this possibility. VIB has an active policy to carefully monitor research data for tech transfer and valorisation potential. If the evaluation indicates that the data requires protection, the invention will be thoroughly assessed and, if needed, the invention will be IP protected (patent protection or copyright protection). The IP protection does not withhold the research data from being made public. In the case a decision of a potentially patent application, it will be planned to avoid any delays in data publication. Specifically, anti-GSEC Nanobodies may be subjected to IP restriction but it will be determined by the generated data, for instance, selectivity of the nanobody and performance to bind epitopes. An expert IP management team pro-actively collects inventions within our institution to asses patentability and protect IP.
Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	☒ Yes ☐ No If yes, please comment: This project is co-funded by the Bright Focus Foundation and an agreement on IP ownership between VIB and this foundation exist.

4. Documentation and metadata	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	Digital data will be generated following standardized protocols and will be stored on VIB/KU Leuven servers. The data will be available at the latest at the time of publication. We do not expect any restraints unless imposed e.g. by pending publications or IP protection of data. Still data and tools may be shared with third parties following a material transfer agreement (MTA). The MTA will clearly define the data and type of reuse allowed. Metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. Metadata will be collected by the research along with the technical stuff at the time of data collection and analysis in the electronic laboratory notebook (E-book) and/or in hard copy notebooks that refer to specific datasets.
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	☐ Yes ☒ No If yes, please specify:

5. Data storage & backup during the FWO project	
Where will the data be stored?	Digital data will be stored on the E-notebook (Electronic Lab Note). The minimum preservation term of 5 years together with the principle of preservation of data will be applied without restriction to raw data as well as processed data. -Vectors: Relevant published vectors and associated sequences will be sent to the non-profit plasmid repository Addgene (which will take care of vector storage and shipping upon request) or be stored in the host institute and shared upon MTA. -Tissue samples: human tissues will be stored in the laboratory and registered with a Belgian biobank in compliance with the Belgian law on human body material (dd 19/12/2008) -Cell lines: Newly created human cell lines will be stored in liquid nitrogen storage and deposited in the UZ Leven-KU Leuven Biobank. Other cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when used for experiments.
How will the data be backed up?	VIB-KU Leuven drives are maintained on a monthly basis, including upgrades and security patches. The servers are maintained by ICTS, and only ICTS personnel (bound by the ICT code of conduct for staff) have administrator rights. Stored data are backed up daily according to snapshot technology, where all incremental changes in respect of previous changes are kept online. The last 14 days backup are maintained. As a consequence, 10% of the requested storage is reserved for backups according to the following scheme: 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 ☐ No If no, please specify: 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.

What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	All digital files will be stored on VIB-KU Leuven servers. The "L-drive" costs 173,78 €/TB/Year. This estimation is based on the costs of digital data storage in the "L-drive" (see the previous point). The "J-drive" costs 519€/TB/Year. This server is based on a cluster of NetApp FAS8040 controllers with an Ontap 9.1P9 operating system. Stored data is backed up using snapshot technology where all incremental changes in respect of the previous version are kept online. Backups are performed hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept. Both servers are accessible only by lab 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. The overall costs will be covered by institutional funding allocated to the host laboratory. The associated costs of storage of published vectors and sequences at non-profit plasmid repositories (Addgene) are negligible. Other vectors generated during the project will be shared with researchers upon request (handling by the technical staff of the laboratory, shipping costs supported by the receiver).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	The “L-drive” server is accessible only by laboratory staff members and is mirrored in ICSTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour. Access to the digital data is possible only using a username-id and a password, and user rights only grant access to the data within their vault. Sensitive data transfer will be performed in line with the best practices for “Copying data to the secure environment” defined by KU Leuven. In the matter of personal and sensitive data, we will conform to the Belgian law of the protection of individuals with regard to the processing of personal data (30/07/2018) and the General Data Protection Regulation 2016/679. The Privacy Team of VIB-KU Leuven will be notified before the start of the project research and will therefore indicate the categories of personnel who have access to sensitive data, this will protect confidentiality. Moreover, all private data will be rendered anonymous before processing outside the digital vault. The PI will be the only one granted access to the server to deposit private data. Only de-identified data will be stored on KU Leuven servers where it can be accessed by staff members of the laboratory.

6. Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.	
Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	All datasets will be stored on the university's central servers with automatic back-up procedure for at least 5 years, conform to the KU Leuven RDM policy. After that, the PI will decide whether retention of data is needed and, if applicable, delete data.
Where will these data be archived (= stored for the long term)?	Long term storage will be ensured as following indicated: - Digital datasets: digital data will be stored on the "L-drive" - Tissue samples: tissues will be stored locally in the laboratory - Vectors: the vectors will be collected in the form of purified DNA - Cell lines: human cell lines will be stored in the UZ Leuven Biobank. In addition, datasets may be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the time of publication.
What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* ***the allocated project budget*** *to be used to cover the cost incurred.*	The total estimated costs for data preservation during the 5 years after the project is ~900 €. This estimation is based on the following costs: - The cost of the digital data storage are the following: 173,78€/TB/Year for the "L-drive" Electricity costs for the -80°C freezers present in the labs are included in general lab costs. Data storage and backup costs are included in general lab costs.

7. Data sharing and reuse	
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?	☒ Yes ☐ No If yes, please specify: The nanobody related work is co-founded by the Bright Focus Foundation and the derived data and IP (aim 1) is covered by an MTA.
Which data will be made available after the end of the project?	Participants of the present study are committed to publish research results communicate them to peers and make them available to the scientific community. All research outputs (materials, protocols, tools, etc) supporting publications will be accessible freely or distributed to other parties upon request and an agreement (MTA) is in place.
Where/how will the data be made available for reuse?	☒ In an Open Access repository ☐ In a restricted access repository ☒ Upon request by mail ☐ Other (specify):
When will the data be made available?	Published datasets will be publicly available as soon as the embargo date is reached, unless IP requirements or ongoing projects require further confidentiality. However, confidential data may be shared to third parties under an MTA.
Who will be able to access the data and under what conditions?	Metadata will contain sufficient information to support data interpretation and reuse and will follow community rooms. Thus, whenever possible datasets and the appropriate metadata will be public available through repositories that support FAIR data sharing.
What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	It is our intent to minimize data management costs by implementing standard procedures. For metadata collection and file storage and organization from the start of the project. Data management costs will be covered by institutional budget allocated to the host laboratory. Moreover, budget for open-access publication costs has been requested to the FWO for this project.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	Research and technical staff will generate, collect, analyse and store the above listed data. Metadata will be documented by taking careful notes in the electronic laboratory notebook (E-notebook) that refers to specific datasets.
Who will be responsible for data storage & back up during the project?	Data storage and back-up will be ensured by the researcher and technical staff. To ensure a correct storage and back-up we will benefit from the support of René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.
Who will be responsible for ensuring data preservation and sharing?	The PI (Lucía Chávez Gutierrez) is responsible for data preservation and sharing, with support from the research and technical staff.
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.