TIMP1 PLAN (FWO DMP)

DMP TITLE

ADMIN DETAILS

Project Name: TIMP1 plan (FWO DMP) - DMP title
Project Identifier: u0149730
Grant Title: 1S11123N - Structure-based design of TIMP1 binding PET tracer for Multiple Sclerosis.
Principal Investigator / Researcher: Ahmed Shemy
Description: Structure-based design of novel TIMP1 binding PET tracer to be used as a tool for clinical evaluation of remyelinating therapies.
Institution: KU Leuven, UHasselt

1. GENERAL INFORMATION

Name applicant
Ahmed Shemy

FWO Project Number & Title
Project number: 1S11123N
Title: Structure-based design of TIMP1 binding PET tracer for Multiple Sclerosis.

Affiliation
- KU Leuven
- UHasselt

2. DATA DESCRIPTION

Will you generate/collect new data and/or make use of existing data?
- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Types of data	Format	Volume	How created
Protocol and SOPs	.xls .docx .pdf	50 MB	LibreOffice Writer, LibreOffice Calc
Purified proteins	1.5 mL – 2 mLtubes	100 tubes	Following protein purification
Glycerol stocks	1.5 mL tubes	100 tubes	Glycerol stock
DNA	1.5 mL tubes	100 tubes	In vitro cloning
Protein structural data (raw xray diffraction images)	.cbf .h5	8 TB	Xray diffraction at synchrotron
Protein structural data (processed)	.pdb .fasta .mtz .cif .mcif .hkl .ASCII .ccp4	20 GB	Coot, Phenix, PyMOL, CCP4i, XDS
MDB files and small molecules databases (pre and post processing)	.mdb .txt	50GB	MOE
PDB files and protein amino acid/DNA sequences	.pdb .fasta .txt .pdf	1GB	MOE, PyMOL, obabel (bash script),Benchling
Scripts	.sh .py	10 MB	Programmed in bash or Python
Chromatograms (SEC, IEX, HIC, Affinity chromatography)	.ASCII .txt .png .jpg	100 GB	Exported from AKTA software
Agarose, SDS, PAGE gelelectrophoresis	.tiff .gel and .jpg	1 GB	Gel images
Spectroscopy data	.txt .png .jpg .ASCII .xls	500 MB	Exported from relevant equipment software (Tecan, JASCO CD spectrometer, Zetasizer Nano DLS, Nanodrop,…), processed
Bio-Layer Interferometry data	.txt .xls .pdf .jpg .png	100 MB	Exported from relevant equipment software (Sartorius Octet R4/R8), processed
Crystal images	.png .jpg	1 GB	Photographed via Nikon microscope
DNA Sequencing	.seq .ab	1 GB	Sequencing performed by company
Analytic UltraCentrifugation graphs	.xls .png .jpg .txt	100 MB	Exported from relevant equipment software andprocessed

3. LEGAL AND ETHICAL ISSUES

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

N.A.

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)

- No

N.A.

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

Small molecules data which are confirmed hits. These compounds will need to be IP protected as they have a potential for valorisation.

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?

- No

N.A.

4. DOCUMENTATION AND METADATA

What documentation will be provided to enable reuse of the data collected/generated in this project?

Each experiment is documented in the (E-)lab journal of the scientist who performs it. Standard operating procedures (SOPs) have been and will be written for all lab techniques used. Notes will describe the biological/clinical samples used, experimental setup and protocols used, data generated, links to the specific location of sample derivatives and data, as well as the names of the datasets. Data obtained from experiments will be stored in specific folders that also contain a README.txt file explaining the design/protocol, analysis methods, results, and labels used in the data analysis file, and a reference to the (E-)lab journal of that particular experiment. This information will allow another researcher to follow all steps in the data processing. Additionally, algorithms, scripts, and software usage will be documented. When scripts, algorithms, and software tools are finalized, they will be further described in manuscripts and/or online Git repositories (e.g. GitHub).

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

- In general each lab will use an electronic lab notebook / paper based lab book in which a number of predetermined topics have to be described for each experiment (objective, protocol, results, and conclusion). Any lab book format used will contain references to the digital data.
- Metadata with the connection between lab samples and files on our data storage so that data files, lab samples, and experimental notes (including descriptions of equipment, setting and used experimental settings) will remain properly linked.
- As a general rule, raw data of every experiment type will be sorted and stored in separate folders. Per work package, processed data will be sorted and stored in separated folders with links to their respective raw data files. Separate documents of non-experimental nature will be sorted and stored in a documents folder.
- Any form of publication can require a combination of data from multiple experiments and work packages. Thus, new folders will be made for all processed data per publication (and links to respective raw data).
- When depositing data in a local or public repository, the final dataset will be accompanied by all relevant information in a README.txt document, following the Dublin Core Metadata standard if no other meta-standard is available yet. 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 and add contextual value to the dataset for future reuse.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

Where will the data be stored?

Throughout the duration of the project, the data will be initially saved on the local workstation and frequently backed up on the group's NAS. At certain milestones or upon completion of the project, the data will be preserved for future use by archiving it on the NAS.

How is backup of the data provided?

Important data in the Voet lab will be saved on a double backup NAS (Synology) that is secure and requires access authorization. Meanwhile, data from the project in other labs will be stored on ICTS-managed, failure-proof systems that have snapshot capabilities.

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 ICTS and VSC provide sufficient storage and archival capacity during and after the project. Additionally, the laboratory of prof. Voet has 55 TB of free space on the NAS system. This should cover all research generated data for the next 4 years.

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

We have estimated a collection of up to 5 TB storage over the duration of the project. At current ICTS/VSC storage pricing of 20-70 Euro/(TB*year) we estimate storage costs at 700-2450 Euro to cover the projects duration and long term storage (10 y). This amount has been budgeted as part of the project.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data is secured by the ICTS service of KU Leuven or is securely stored on a Synology NAS system. Confidential data can and will be protected with a password, available only for PIs Arnout Voet and Niels Hellings. Visitors, MSc thesis students and internship students in the groups as well as other unauthorized persons will not have access to the data on the shared folder. The VSC storage is only accessible to VSC accounts, and the respective volumes will only be accessible to group members.

6. DATA PRESERVATION AFTER THE FWO PROJECT

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, ...).

The data to be retained during 10 years after the project's end are dissemination data (source files of publications, presentations, and patents) and the most relevant measurement data. A possible exception to this could become the raw imaging data files from the xray diffraction experiments, but only if it will be judged that keeping the reconstructed/processed imaging data is not necessary and the cost of storing the large amount of raw imaging data will be unaffordable.

Where will the data be archived (= stored for the longer term)?

The research data will be stored in a glacial archive storage system after the end of the project. Dissemination data, namely files corresponding to papers and presentations, will be stored on the PCs of involved PIs, and backed-up daily on the departmental server for long term storage. Analysis code will be stored on ICTS-hosted code repositories.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

- The volume corresponding to dissemination data is expected to be relatively low, and therefore can be seamlessly embedded in the PIs' allocation on the departmental server.
- For research data, at current archiving costs of 10 Euro/(TB*year), we estimate a cost of 2000 Euro/year. These costs will be covered by funding acquired by the project PIs in the context of other research projects.

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)?

- No

N.A.

Which data will be made available after the end of the project?

Relevant digital data will be published and made available after the end of the project. Data with valuable IP will be protected prior to publication. We will comply with open access regulations of KU Leuven.

Where/how will the data be made available for reuse?

- Upon request by mail

The approach to share data is upon request by e-mail. Due to the data volume, access will then be granted to a restricted access repository.

When will the data be made available?

- Upon publication of the research results

As soon as the research results have been published, the data can be made available to other researchers.

Who will be able to access the data and under what conditions?

All project collaborators will be authorized to have access to all obtained digital and physical data after the project. In case the question originates by researchers outside the consortium, the data can be made available upon e-mail request, and on condition that the users agree to give proper credit, such as co-authorship on their papers building on these data. Usage for commercial purposes will require obtaining a license, or equivalent arrangement.

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

A restricted access repository can be implemented on a free tool, such as Dropbox, up to a certain volume. If this volume does not suffice, time-limited storage will be considered, thus limited to the time needed to download the data. Furthermore, the laboratory of prof. Voet stores essential data in a 96 TB NAS drive with secure cloud-based backup and online access via Synology. Later, all datasets will be made public via relevant databases (eg. PDB, GitHub repositories, Mendeley data repositories).

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?
The researcher will be responsible for data documentation & metadata.

Who will be responsible for data storage & back up during the project?
The researcher will be responsible for data storage & back up during the project.

Who will be responsible for ensuring data preservation and reuse ?
The researcher, during the course of this project, and PI, after completion of this project, bear the end responsibility of ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?
The PI bears the end responsibility of updating & implementing this DMP.