DMP title

Project Name DMP for Junior postdoctoral fellowship of FWO - DMP title

Grant Title 1229222N

Principal Investigator / Researcher Hans Gerstmans

Description DMP for Junior postdoctoral fellowship of FWO. Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a rapidly expanding class of chemically and structurally diverse bacterial natural products. Many exhibit therapeutically important biological activities, such as antibiotic, antifungal, insecticidal, immunomodulating and anti-cancer activities. The relaxed substrate specificity of RIPP biosynthetic enzymes, together with the increasing availability of methods to efficiently introduce multiple site-directed mutations, opens the door to generating large libraries of RiPP variants that can be screened for improved pharmaceutical properties. However, there is currently no suitable screening platform that can handle such enormous amounts of variants. We envision that a droplet-based microfluidic platform is the ideal technology to handle high numbers of engineered RiPP variants. A microfluidic droplet is the smallest bioreactor imaginable; it is the ultimate miniaturization of a microtiter plate well. This allows for ultrahigh-throughput parallelization of assays with an analysis rate of up to 2000 variants per second. In this project, we therefore aim to develop a first in its kind droplet-based microfluidic platform to screen libraries of engineered RiPPs for variants with new biological activities that can be used for therapeutic applications.

Institution KU Leuven

1. General Information

Name applicant

Hans Gerstmans

FWO Project Number & Title

1229222N: A microfluidic platform to screen engineered lasso peptides for therapeutic potential

Affiliation

- KU Leuven
- Other

Vlaams Instituut voor Biotechnologie (VIB)

The promotor of this project is prof. Joleen Masschelein. Prof. Jeroen Lammertyn is co-promotor.

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).

Generate new data

Name	Origin of data	Type of data	File format	Volume
Mutant bacterial strains	DNA engineering and deletion experiments	Stored in duplicate as glycerol stocks in a -80°C freezer	Inventory and descriptions in excel file (xlsx)	1 GB
Plasmid constructs	Protein purification and gene deletions experiments	Stored in duplicate in a -20°C freezer	Inventory and descriptions in excel file (xlsx)	1 GB

DNA libraries of RiPPs	DNA engineering experiments	Stored in duplicate in a -20°C freezer	Inventory and descriptions in excel file (xlsx)	10 GB
DNA agarose electrophoresis gels	DNA engineering and deletion experiments	Digital images	TIFF	1 GB
DNA sequencing of plasmids	Confirm DNA engineering experiments by Sanger sequencing	Sequencing files	Orginal: ABI, inventory and descriptions of results in excel file (xlsx)	1 GB
Complete plasmid and genome sequencing	Confirm sequence of new plasmid and bacteria	Sequencing files	FASTQ	100 GB
SDS PAGE gels	Expression and purification of engineered RiPP variants	Digital images	TIFF	1 GB
LC-MS data	LC-MS analyses of bacterial metabolites and enzymatic reactions	LC-MS chromatograms	FID	5 TB
HPLC data	HPLC purification of engineered RiPP variants	HPLC chromatograms	PDF	2 TB
NMR data	NMR spectroscopic structural analyses of engineered RiPP variants	NMR spectra	FID	10 GB
MIC and MBC data	Antibacterial spectrum and potency of engineered RiPP variants	Pictures and spectroscopic data (absorbance)	TIFF and descriptions of results in excel file (xlsx)	5 GB
OrganoPlate® system	Evaluating engineered RiPP variants in gut-on-a-chip model and cytotoxicity testing	Pictures, spectroscopic data (bright field, fluorescence)	TIFF and descriptions of results in excel file (xlsx)	100 GB
Designs of microfluidic chips	Designs of microfluidic chips	Designs	CAD	500 GB

Validation and running of microfluidic chips	Validation and running of the microfluidic chips to screen engineered RiPP variants	Pictures and spectroscopic data (brightfield, fluorescence)	TIFF, csv, matlab	1 TB
Lab book notes and reports	Description of experiments and results	Text and pictures	word	5 GB

Reuse existing data

Name	Origin of data	Tools used to analyse data	File format	Volume
Finding potential new RiPP tailoring enzymes	Public available genomes and protein sequences	RODEO, antiSMASH, RiPPMiner, RiPPER and EFI-EST	CSV, GBK, txt	500 MB

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

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)

- No

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

Potential engineered RiPP variants with improved antibacterial properties resulting from this project will be protected through patent applications which will enable future exploitation and commercial development. This will be done in collaboration with both LRD (the University’s office for intellectual property and technology transfer) and VIB ventures.

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

4. Documentation and metadata

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

1. Microscopy images the following information will be noted: dimensions, image type, bit-depth, pixel sizes and microscope settings. A ReadMe file of the image collection will be written.
2. Raw data will be collected per test, including a txt file with a clear description of what the data represent and how they were generated. The input files used for the simulation will be kept in the same folder. The name of the folder will contain the composition, temperature and a

reference to the loading conditions of the considered material (A .txt file explaining the naming will be maintained).

The methodology, protocols, dates and results for all experimental procedures will be documented in detail in electronic lab books (including link with physical storage of data, .doc files). Every month, a time-stamped pdf copy of each lab book will made and stored on the KU Leuven servers. General protocols and standard operating procedures will be collected in a dedicated folder on a network drive (J:) at KU Leuven. A detailed inventory of the bacterial strains and DNA plasmid generated thoughout the course of this project will also be stored on this network drive in the form of excel files.

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.

- No

No metadata standard is available for the type of data that we will generate. The metadata for all NMR, LC-MS and UV spectroscopic data will be available within the file format that they will be stored in. The metadata for all other experimental procedures (e.g. setup, preparation, date, storage, ...) will be generated by the researchers in a descriptive format that enables easy interpretation by other people in the future.

5. Data storage and backup during the FWO project

Where will the data be stored?

Data will be stored on a network drive (J:) which is backed-up by the ICTS service of KU Leuven. This includes monthly, time-stamped copies of each lab notebook, images and spectroscopic data. Additional copies will be made and kept on personal devices. The network drive is only accessible to group members. Once a researcher leaves the lab, their data will be transferred to a large volume network archive drive. All LC-MS data will be stored in duplicate on dedicated external hard drives. Bacterial strains and DNA plasmids will be kept in duplicate in -80 and -20°C freezers, respectively.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures. Copies of the LC-MS data will be stored on dedicated external hard drives.

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

Yes, KU Leuven provides sufficient storage and back-up capacity during and after the project. Dedicated external hard drives of 5 and 50 Tb are available for storing the LCMS data.

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

The KU Leuven network drive costs 51.9 % per 100 Gb, the large volume costs 113€ per Tb. These costs, along with the costs for the external hard drives (approx. 200 €) will be covered by the project budget.

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

All data on the KU Leuven network drives is only accessible by the project researchers. This is determined by their KU Leuven personnel number. Furthermore, the data on the external hard drives will be protected by a password.

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, ...).

All data obtained during this FWO project will be retained for at least the expected 5 year period after the end of the project.

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

The data will be stored indefinitely on the university's large volume network archive drive (with automatic back-up procedures). The LC-MS data will be stored in duplicate on external hard drives.

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

The KU Leuven large-volume network archive drive costs 113€ per Tb. These costs, along with the costs for the external hard drives (approx. 200 €) will be covered by the project budget.

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. Specify:

If novel engineered RiPP variants with improved properties are found or generated, a patent application will be filed. This may temporarily restrict the sharing of data.

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

All published data will be made available. Data with valuable IP will be protected prior to publication.

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

- In an Open Access repository
- Upon request by mail

The data will be stored and be available for lab members using a shared network drive. Published data will be available for everyone.

When will the data be made available?

- Upon publication of the research results

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

Only researchers participating in the project and lab members will be able to access the data before
publishing. Upon publication, everyone will be able to access the data.

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

None.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The researcher working on this project (Hans Gerstmans) and the PI's (Joleen Masschelein and Jeroen Lammertyn) will be responsible for data documentation & metadata.

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

The researcher working on this project (Hans Gerstmans) and the PI's (Joleen Masschelein and Jeroen Lammertyn) will be responsible for data storage & backup during the project.

Who will be responsible for ensuring data preservation and reuse ?

The researchers working on this project (Hans Gerstmans) and the PI's (Joleen Masschelein and Jeroen Lammertyn) will be responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

The PI's (Joleen Masschelein and Jeroen Lammertyn) bear the end responsibility of updating & implementing this DMP.