DMP for FWO-SB fellowship - RiPPing through pathogenic bacteria and protein-protein interactions

A Data Management Plan created using DMPonline.be

Creators: Hanne Vande Capelle, Joleen Masschelein, Hans Gerstmans

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Joleen Masschelein

Grant number / **URL:** 1S28323N

ID: 194331

Start date: 01-10-2022

End date: 31-10-2026

Project abstract:

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. Sactipeptides are macrocyclic RiPPs featuring a unique hairpin-like structure that provides a promising scaffold for the development of novel therapeutics. By combining the promiscuity of RiPP biosynthesis with a droplet-based microfluidic screening platform, I aim to push the boundaries regarding RiPP engineering for therapeutic applications. In this project, I will construct and screen unique sactipeptide libraries for novel derivatives with activity against oncogenic gut pathogens as well as selective protein-protein interaction inhibitors.

Last modified: 23-04-2023

DMP for FWO-SB fellowship - RiPPing through pathogenic bacteria and protein-protein interactions

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	
Plasmid constructs	Plasmid constructs are generated to enable the recombinant production of proteins in bacteria. The plasmids are stored in duplicate in a -20°C freezer. An inventory and descriptions are stored in an Excel file.	Generate new data	Physical (plasmids) and digital (inventory)	Metadata: Excel file with inventory and description of plasmids	.xlsx	< 100 MB	Some boxes of 1.5 mL Eppendorf tubes
Sequencing data Sanger sequencing	DNA is sequenced to confirm DNA engineering. This is done by sending the DNA samples for Sanger sequencing to Eurofins Genomics (Night Xpress). They send the data back as .abi file to process. The data is processed in SnapGene and stored as .dna file. An inventory and descriptions are stored in an Excel file.	Generate new data	Digital	- Other: DNA sequencing from an external company - Metadata: excel file with inventory and description of sequencing data	- Received as .abi and stored as .dna. - .xlsx	< 100 GB < 100 MB	
Sequencing data complete plasmid and genome sequencing	Confirm the sequence of new plasmids and bacteria via Illumina and Nanopore sequencing.	Generate new data	Digital	- Other: DNA sequencing - Metadata: excel file with inventory and description of sequencing data	- .fastq - .xlsx	< 100 GB	
Finding potential new RiPP tailoring enzymes by bioinformatic analysis	Search in publicly available genome and protein sequences. Bioinformatic tools like RODEO, antiSMASH, RiPPMiner, RiPPER, EFI-EST are used for this purpose.	Reuse existing data	Digital	Software	CSV GBK txt	< 1 GB	
Mutant bacterial strains	Mutations are introduced in the genome of bacterial strains. These experiments make the bacteria more suitable for heterologous expression and screening purposes. The bacteria are stored in duplicate as glycerol stocks in a -80°C freezer. An inventory and descriptions are stored in an Excel file.	Generate new data	Physical (bacteria) and digital (inventory)	Metadata: Excel file with inventory and description of plasmids	.xlsx	< 100 MB	Some boxes of 2 mL glycerol stocks
DNA libraries of sactipeptide	Libraries of sactipeptide variants. The libraries are stored in duplicate in a -20°C freezer. An inventory and descriptions are stored in an excel file.	Generate new data	Physical (libraries) and digital (inventory)	Metadata: excel file with inventory and description of plasmids	.xlsx	< 100 MB	Some boxes of 1.5 mL Eppendorf tubes
DNA agarose electrophoresis gels	To visualize DNA. To verify whether a PCR reaction succeeded or whether DNA is present. A digital image of the gel is made to store.	Generate new data	Digital	experimental/observational	.TIFF .jpg .png	< 1 GB	
SDS page gel	To visualize proteins to verify whether a protein is produced. A digital image of the gel is made to store.	Generate new data	Digital	experimental/observational	.TIFF .jpg .png	< 1 GB	
Underlay and overlay assays for activity testing	To give a first indication on whether the sactipeptide is produced. The size of the halos gives an indication of the amount produced.	Generate new data	Digital	experimental/observational	.TIFF .jpg .png	< 1 GB	
MIC and MCB data	To look at the antibacterial spectrum and potency of the sactipeptide variants. The absorbance is measured and stored in an Excel file.	Generate new data	Digital	experimental/observational	.TIFF .jpg .png .xlsx	< 1 GB	

Miller assay	To test the strength of protein-protein interactions in a liquid medium. The absorbance will be measured and stored in an Excel file. This file will also be used for calculations of the Miller coëfficient.	Generate new data	Digital	experimental/observational	.xlsx	< 100 MB	
LC-MS data	LC-MS analysis is performed to verify the presence, molecular weight and quantity of the compounds of interest.	Generate new data	Digital	experimental: LC-MS chromatograms	.FID	< 5 TB	
HPLC data	HPLC purification of the engineered sactipeptide variants.	Generate new data	Digital	experimental: HPLC chromatograms	.PDF	< 1 TB	
NMR data	NMR spectroscopic structural analysis of engineered sactipeptide variants	Generate new data	Digital	experimental: NMR spectra	.FID	< 100 GB	
Lab book notes and reports	Description of all of the experiments performed in the lab and the results obtained	Generate new data	Digital	Compiled/aggregated data	.docx	< 1 GB	
Design of microfluidic chips	Design of microfluidic chips for the high-throughput screening of sactipeptide variants	Generate new data	Digital	Others: designs	.CAD	< 1 TB	
Validation and running of microfluidic chips	Validation and running of microfluidic chips for high-throughput screening of sactipeptide variants	Generate new data	Digital	Others: Pictures and spectroscopic data (e.g. brightfield, fluorescence)	.TIFF .csv .matlab	< 100 GB	

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:

RODEO: https://ripp.rodeo/advanced.html
antiSMASH: https://antismash.secondarymetabolites.org/#!/start
RiPPMiner: http://www.nii.ac.in/~priyesh/lantipepDB/new_predictions/index.php
RiPPER: https://github.com/streptomyces/ripper
EFI-EST: https://efi.igb.illinois.edu/efi-est/

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.

- No

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

Potential engineered sactipeptides with activity against colorectal cancer-associated bacteria or with inhibitory activity of protein-protein interactions resulting from this project will be protected through patent applications which will enable future exploitation and development. This will be achieved in collaboration with both LRD (the KU Leuven university's office for intellectual property and technology transfer) and VIB ventures. Besides, a heterologous expression platform for the production of sactipeptide variants resulting from this project will be protected through patent applications in the same way.

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).

Raw data will be collected per test, including a .txt file with a clear description of what the data represents and how they were generated. An inventory of all data and a description will be collected in an Excel file. All work performed in the lab will be saved as an Electronic Lab Notebook (ELN). This notebook will contain all information on methodology, protocols, results and conclusions. The ELN will be ordered chronologically, with a title for every date and a subtitle for every experiment. Every month, a time-stamped pdf copy of the ELN of that month will be made and stored on the KU Leuven servers. General protocols and standard operating procedures will be collected in a dedicated folder on a Shared network drive (J:) at KU Leuven.

All the images with their explanation will be saved in the ELN. Additionally, the images will be saved in separate folders. The name of this folder consists of the date and the title of the experiment.
Various tools will be used to process data (e.g. SnapGene for sequencing data ...). The input files will be kept in the same folder as the processed files. The name of the folder will contain the date and information about the experiment. Background information will be stored in an Excel file and the Electronic Lab Notebook.

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.

- No

No metadata standard is available for the type of data that will be generated. The metadata for all NMR, LC-MS and UV spectroscopic data will be available within the file format they will be stored in. The metadata for all other experiments will be generated in a descriptive format which can be interpreted easily by other people in the future.

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

Where will the data be stored?

All data will be transcribed into digital format and will be stored in an online lab management system with an integrated Electronic Lab Notebook together with accompanying information. This Electronic Lab Notebook (monthly time-stamped copies) and digital data (e.g. images, spectroscopic data, sequencing data, metadata ...) will be stored in a personal folder on a Shared network drive (J:) which is backed up by the ICTS service of the KU Leuven. Additional copies will be made and kept on personal devices. Once a researcher leaves the lab, their data will be transferred to a large-volume network archive drive. Besides, the LC-MS data will be stored in duplicate on dedicated external hard drives. Bacteria and DNA will be stored in duplicate in -80 and -20°C freezers, respectively.

How will the data be backed up?

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

KU Leuven provides sufficient storage and backup capacity during and after the project. The initial storage capacity of the server of KULeuven is 5 GB. However, it can be extended without extra cost to 10 GB. Besides, dedicated external hard drives of 5 and 10 TB are available for storing the LC-MS data.

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

The Shared network drive (J:) of the KU Leuven is only accessible to group members. Their access is determined by their KU Leuven personnel number. The drive has a high level of security. Furthermore, the data on the external hard drives will be protected by a password.

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

The KU Leuven shared network drive costs € 503,66/ TB / year. Additionally, a large volume storage drive from the KU Leuven costs € 104,42 / TB / year. This drive will be used to store all the large files. These costs, along with the costs for the external hard drives (approx. € 200) will be covered by the project budget.

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...).

After the research project, the principal investigator (Prof. Joleen Masschelein) will take responsibility for data preservation. All the data will be stored in an online lab management system with an integrated Electronic Lab Notebook. The data will be preserved for 10 years according to KU Leuven RDM policy.

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

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

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

The KU Leuven large-volume network archive drive costs € 104,42 / TB / year. These costs and the costs of the external hard drives (approximately € 200) will be covered by the project budget.

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 a restricted access repository (after approval, institutional access only, …)

During the project, all data will be stored on the shared network drive of the KU Leuven. Lab members and people participating in the project can get access to the data stored on this drive based on their personnel number.
If novel engineered sactipeptide variants with improved properties are found or generated, a patent application will be filed. This may temporarily restrict the sharing of data.
After the end of the project, all published data will be made available.

If access is restricted, please specify 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.

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

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

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

After publishing, the data will be made available in Lirias, the open access repository of the KU Leuven. Genome data will be made available in GenBank (NCBI).

When will the data be made available?

The data will be made available upon publication of the research results.

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

If novel engineered sactipeptide variants with improved properties are found or generated, a patent application will be filed. Data can be accessed and reused upon request by email.

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

A unique identifier will be added to the published data. Besides, genome data deposited in GenBank will also get a unique accession number.

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

There are no expected costs for data sharing.

6. Responsibilities

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

During the research project, the fellowship holder (Hanne Vande Capelle), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Hans Gerstmans) will manage the data documentation and metadata.

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

During the research project, the fellowship holder (Hanne Vande Capelle), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Hans Gerstmans) will manage data storage and backup.

Who will manage data preservation and sharing?

During the research project, the fellowship holder (Hanne Vande Capelle), the Principal Investigator (Prof. Joleen Masschelein) and the co-promoter (Dr. Hans Gerstmans) will manage data preservation and sharing.

Who will update and implement this DMP?

This data management plan will be updated and implemented by the fellowship holder (Hanne Vande Capelle). The Principal Investigator (Prof. Joleen Masschelein) bears the end responsibility of updating and implementing this DMP.