Charting the biosynthetic landscape of a novel class of bioactive specialized lipids

A Data Management Plan created using DMPonline.be

Creators: Fatima El Arnouki Belhaji, Joleen Masschelein, Angus Weir

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: 11PP324N

ID: 204370

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Microorganisms produce a wealth of specialized metabolites with highly diverse chemical structures and important applications in medicine and agriculture. The zeamines are an unusual group of longchain polyamine antibiotics produced by Serratia bacteria that exhibit potent activity against a broad spectrum of (micro)organisms, including bacteria, fungi, plants and nematodes. They are assembled by a unprecedented combination of polyketide, nonribosomal peptide and polyunsaturated fatty acid synth(et)ase-like biosynthetic machinery. A preliminary bioinformatic search has indicated that such hybrid biosynthetic pathways are not an isolated occurrence, but are in fact an underexplored phenomenon. The zeamines are therefore proposed to represent the first identified members of an entirely new class of biologically active, hybrid specialized lipid metabolites that have remained elusive so far. This project aims to gain a comprehensive understanding of the distribution and diversity of these exceptional hybrid specialized lipid biosynthetic pathways in bacteria. It also aims to expand the catalogue of known specialized lipid-containing metabolites and investigate their bioactivity, biosynthesis and ecological role. The outcomes of this project will not only provide novel bioactive metabolites with unique chemical structures, but it will also shed light on a new biosynthetic paradigm and the molecular mechanisms that enable this level of functional crosstalk.

Last modified: 22-03-2024

Charting the biosynthetic landscape of a novel class of bioactive specialized lipids
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, .csv,.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 will be designed to generate Knock-Out (KO) mutants, and for the recombinant production of proteins in bacteria. These plasmids will be stored at -20°C, at least in duplicate, and an inventory and description will be included in an excel file.	Generate new data	Physical (plasmids) and digital (repository)	Metadata: Excel file with inventory and description of the plasmids.	.xlsx	< 100 MB	Boxes with 1.5 mL Eppendorf tubes.

Sanger sequencing data	Sanger sequencing of DNA confirms engineering of DNA molecules. DNA samples are sent to Eurofins Genomics (Night Xpress) for Sanger sequencing. The results are provided as .abi files. These are further processed with SnapGene, generating .dna files. An inventory and description of all data is saved in an Excel file.	Generate new data	Digital	• Other: DNA sequencing data (external provider) • Metadata: Excel file with inventory and description of the DNA sequencing results.	• Received as .abi and stored as .dna • .xlsk	• < 100 GB • < 100 MB	
Complete genome and plasmid sequencing data	Obtention of the full DNA sequence of a genome or a plasmid via Illumina and Nanopore sequencing	Generate new data	Digital	• Other: DNA sequencing data (external provider) • Metadata: Excel file with inventory and description of the sequencing data	• .fastq • .xlsx	< 100 GB	

Mutant bacterial strains	Introduction of mutations in the DNA of bacteria for different purposes (knocking-out a gene, changing the level of expression of an operon, etc.). Glycerol stocks of the mutated bacteria are kept at -80°C, at least in duplicate. An Excel file is generated with inventory and description of the strains. One mutant of the species of Paraburkholderia megapolitana was previously generated by Jelle Dillen during his Master Thesis under the supervision of Rob Lavigne and Joleen Masschelein (KU Leuven, 2019)	Generate new data + Reuse existing data (mutant by Jelle Dillen; KU Leuven, 2019)	Physical (bacteria) and digital (inventory)	Metadata: Excel file with inventory and description of mutated bacteria strains.	.xlsx	< 100 MB	Boxes with glycerol stock tubes with the mutated bacteria.
Finding new hybrid lipid biosynthetic pathways	Looking for new hybrid lipid biosynthetic pathways which are similar to the zeamine pathway. Bionformatic tools like Clustertools, antiSMASH and CAGECAT, among others will be used. The search could be based on conditions established in collaboration with Emmanuel de Los Santos (Warwick University, 2018)	Generate new data + Reuse existing data (Emmanuel de Los Santos, 2018)	Digital	Software	CSV, txt	< 100 GB	

DNA agarose electrophoresis gel	Visualization of results of PCR reactions. A digital image is generated and saved.	Generate new data	Digital	Experimental/observational	.TIFF .jpg .png	< 1 GB	
SDS-PAGE electrophoresis gel	Visualization of proteins. A digital image is taken and saved.	Generate new data	Digital	Experimental/observational	.TIFF .jpg .png	< 1 GB	
LC/MS data	LC/MS analysis is used to monitor enzymatic assays, the presence of bacterial proteins or metabolites, as well as their molecular weight, abundance and purity.	Generate new data	Digital	Experimental: LC/MS chromatograms and spectra	.FID	< 5 TB	
HPLC data	HPLC purification of bacterial metabolites of interest.	Generate new data	Digital	Experimental: HPLC chromatograms	.PDF	< 1 TB	
NMR data	Elucidation of the structure of hybrid lipid metabolites	Generate new data	Digital	Experimental: NMR spectra	.FID	< 100 GB	
Lab book and reports	Description of executed experiments and obtained results	Generate new data	Digital	Compiled/aggregated data	.docx	< 1 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:

antiSMASH: https://antismash.secondarymetabolites.org/#!/start

MIBiG: https://mibig.secondarymetabolites.org/

Paraburkholderia megapolitana mutant of zeamine-like BGC, generated by Jelle Dillen, during his Master Thesis in 2019 at KU Leuven university under the supervision of Rob Lavigne and Joleen Masschelein.

Rules and results obtained in collaboration with Emmanuel de los Santos (Warwick University) for identification of novel hybrid lipid zeamine-like pathways (2018).

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

If zeamine-like compounds with interesting bioactivies are identified in this project, these will be protected through a patent which will enable future exploitation and commercial 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.

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).

The methodology, protocols, and results of all experiments will be described in an electronic lab book (.doc file). Each month, a time-stamped copy of each lab book is generated and stored on KU Leuven servers. New protocols will be stored on a specific folder on a network drive (J:) at KU Leuven. Bacterial strains (WT and generated KOs) and DNA plasmids will be described in an excel file that will also be stored on the network drive.
Raw LC/MS data will be collected alongside a .txt file with description of the experimental conditions.

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 created. The metadata for NMR and LC/MS data will be available within the file format they are saved in. The metadata of remaining experiments will be generated in a format easily interpreted by other people in the future.

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

Where will the data be stored?

All data (time-stamped copies of lab notebooks, images and other digital data) will be stored in a personal folder on a shared network drive (J:) backed up by the ICTS service of the KU Leuven university. Additional copies will be made and stored on personal devices. Once the researcher leaves the lab, their data will be transferred to a large-volume network archive drive. In addition, 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 both during and after the project. Dedicated external hard drives of 5 and 50 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 network drive (:J) is only accessible to researchers working on the project. This is determined by their KU Leuven personnel number. The drive has a high level of security. The data on external physical 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, while a large volume storage drive from the KU Leuven costs € 104,42 / TB / year. These costs, alongside the ones 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...).

All data obtained during the project will be retained and stored, for a total of 10 years according to KU Leuven RDM regulations.

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, the data will be stored on the shared network drive, for which access is restricted to researchers involved in the project. If novel hybrid lipid metabolites with interesting bioactivities are identified, an application for a patent will be carried out. As a result, the sharing of some data may be restricted temporarily. At 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 involved in the project and lab members will have access before publication. Published data will be available for everyone.

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 hybrid lipid metabolites with interesting bioactivities are identified, a patent will be applied for. As a result, the sharing of some data might be restricted temporarily.

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

The data will be available via Lirias, the open access repository of the KU Leuven.

When will the data be made available?

After publication of the results.

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

If novel hybrid lipid metabolites with interesting bioactivities are identified, a patent will be applied for. 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

Yes, a unique identifier will be added.

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

None.

6. Responsibilities

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

The fellowship holder (Fatima El Arnouki Belhaji), and the Principal Investigator (Joleen Masschelein) will manage data documentation and metadata.

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

The fellowship holder (Fatima El Arnouki Belhaji), and the Principal Investigator (Joleen Masschelein) will manage data storage and backup.

Who will manage data preservation and sharing?

The fellowship holder (Fatima El Arnouki Belhaji), and the Principal Investigator (Joleen Masschelein) will manage data preservation and sharing.

Who will update and implement this DMP?

This DMP will be updated and implemented by the fellowship holder (Fatima El Arnouki Belhaji). The Principal Investigator (Prof. Joleen Masschelein) bears the end responsibility of updating and implementing this DMP.