Protein Aggregates: A Window into Bacterial Antibiotic Tolerance and Dormancy

A Data Management Plan created using DMPonline.be

Creators: Natalie Verstraeten, Jan Michiels

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Jan Michiels

Grant number / **URL:** G033324N

ID: 207113

Start date: 01-01-2024

End date: 31-12-2027

Project abstract:

Millions of people worldwide suffer from long-term bacterial infections that cannot be cured by antibiotic treatment. Antibiotics are often incapable of killing all bacteria in a chronic infection, leaving behind a reservoir of cells from which relapse can occur. Increasing evidence shows that some bacteria survive treatment without the development of resistance. These bacteria, named persisters, are reprogrammed to a temporary antibiotic-tolerant state. Recent findings demonstrate that persister formation is tightly linked to, and may be driven by, protein aggregation. Moreover, we have shown the existence of different dormancy depths with extensive aggregation causing a deeper state of dormancy from which bacteria cannot recover. In this respect, persisters are considered shallowly dormant cells that can exit their dormant state to re-establish an infection once the antibiotic treatment ceases. Interestingly, reversion to the non-persister state and subsequent regrowth seems to occur concomitantly with disaggregation. However, a postulated causal relationship between protein aggregation and dormancy remains poorly understood. In this project we will study the dynamics of aggregate composition and structure in relation to bacterial dormancy depth and regrowth. Our results will strongly contribute towards a better understanding of cellular dormancy and how this physiological state enables bacteria to become tolerant to antibiotic treatment.

Last modified: 13-05-2024

Protein Aggregates: A Window into Bacterial Antibiotic Tolerance and Dormancy
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Protein Aggregates: A Window into Bacterial Antibiotic Tolerance and Dormancy
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Protein Aggregates: A Window into Bacterial Antibiotic Tolerance and Dormancy Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

- **Biological material**: *E. coli* MG1655 wild-type, pooled CRISPRi libraries and mutants constructed through genome engineering.
- **Experimental results**: proteomics data, RNA-seq data, CRISPRi data, results from phenotypic analyses including persister parameters, microscopy data, flow cytometry data, results from intracellular infection assays, and other data types that are unknown at this point.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

- **Responsible person**: Jan Michiels.
- **Biological material**: Cryotubes and multi-well plates will be stored in -80° freezers with restricted access (costs covered by general lab expenses).
- **Experimental results**: Data will be stored on secure university servers with built-in backup and versioning; password-protected hard drives equipped with anti-virus programs will be used as backup (costs partly covered by this project).

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

- **Biological material**: All strains will be stored for at least 5 more years after the end of the project. For this purpose, -80°C freezers are present in the Michiels lab.
- **Experimental results**: After the project, all data will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs will be covered by KU Leuven overhead budgets.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Not applicable.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

- **IP**: Ownership of the generated data belongs to KU Leuven; copyright of the data belongs to Jan Michiels.
- **IP 3rd party**: There is no IP on the strains preventing us from storing them, performing anticipated experiments or publishing results.
- **Documentation**: Data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in the same folder also containing the results.

Protein Aggregates: A Window into Bacterial Antibiotic Tolerance and Dormancy
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.

Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
T1.1. Protein content of aggregates	Composition of protein aggregates determined by re-solubilizing isolated aggregates, subjecting them to trypsin digestion and analyzing their composition by MS	New	Digital	MS data	.xlsx, .csv	<1GB	/
T1.1. Protein localization in aggregates	Microscopic evaluation of the location of identified proteins in aggregates	New	Digital	Microscopy images	.jpeg, .tiff, .nd2	<100GB	/
T1.2. mRNA content of aggregates	mRNA content of isolated aggregates determined by transcriptome analyses	New	Digital	RNA-seq data	.txt	<100GB	/
WP1 *E. coli* mutants	CRISPRi or knockout strains with reduced levels of key aggregation proteins and mRNAs	New	Physical	/	/	/	~100 strains, stored in 96 well plates at -80°
WP1 Phenotypic analysis of constructed mutants	Assessment of dormancy following established protocols	New	Digital	Growth curves, flow cytometry data	.xlsx, .xit	<1GB	/
T2.1. LLPS-like state of aggregates	Microscopic evaluation of aggregates (time-lapse epifluorescence microscopy with and without 1,6-hexanediol, FRAP, PALM)	New	Digital	Microscopy images	.jpeg, .tiff, .pzfx, .nd2	<100GB	/
T2.2. Aggregate structure	AFM-IR analysis to obtain three-dimensional morphology, infrared absorption spectra and stiffness maps of aggregates	New	Digital	AFM-IR spectra	.svg, .pzfx	<100GB	/
T3.1 *E. coli* polyP mutants	Low/no and high polyP producers for the 12 strains with contrasting aggregation and dormancy phenotypes described in T1.1	New	Physical	/	/	/	~100 strains, stored in 96 well plates at -80°
T3.1 Phenotypic analysis of constructed strains and wild types treated with a polyP inhibitor	Assessment of aggregate formation, persisters, CFUs and VBNCs	New	Digital	Growth curves, flow cytometry data	.xlsx, .xit	<1GB	/
T3.1 Quantification of polyP levels	Detection of polyP at single-cell level using a fluorescent reporter	New	Digital	Population and single-cell level (flow cytometry)	.xlsx, .xit	<1GB	/

T3.2 Role of polyP in fluidity state of aggregates	Microscopic evaluation of aggregates (time-lapse epifluorescence microscopy with and without 1,6-hexanediol, FRAP, AFM-IR)	New	Digital	Microscopy images, AFM-IR spectra	.jpeg, .tiff, .nd2, .pzfx, .svg	<100GB	/
T4.1 *E. coli* strains tagged with fluorescent disaggregation proteins	Mutants selected in T1.1 and *dnaK* and *clpB* mutants that can no longer dissolve aggregates	New	Physical	/	/	/	~100 strains, stored in 96 well plates at -80°
T4.1 *E. coli* strains expressing fluorescent proteins (essential + non-essential) present in aggregates	Hits selected in T1.1 fluorescently labeled and expressed from an inducible promoter	New	Physical	/	/	/	~100 strains, stored in 96 well plates at -80°
T4.1 Mother machine data	Microscopic monitoring of disaggregation, protein localization and cell division using the mother machine	New	Digital	Microscopy images	.jpeg, .tiff, .nd2	<100GB	/
T4.2 MG1655 sgRNA library	sgRNA library targeting each protein-coding gene as well as ~10,000 intergenic regions of MG1655 (Nat Commun (2018) 9:2475)	Reused	Physical	/	/	/	~20 cryotubes stored at -80°C
T4.2 CRISPRi screening data	Identification of proteins involved in disaggregation using CRISPRi	New	Digital	sgRNAs counts following growth of clones in selective conditions	sam, bam, .ab1, .fasta/fa, .qual, gb/gbk, .dna	<1G	/
T4.2 *E. coli* mutants expressing hit sgRNAs	Validation of CRISPRi results	New	Physical	/	/	/	~100 strains, stored in 96 well plates at -80°
T4.2 Phenotypic analysis of constructed strains	Assessment of aggregate formation, persisters, CFUs and VBNCs	New	Digital	Growth curves, flow cytometry data	.xlsx, .xit	<1GB	/
T5.1 Cellular infection model	Assessment of the development and presence of aggregates in J774A.1 cells	New	Digital	Microscopy images, growth curves, flow cytometry data	.jpeg, .tiff, .nd2, .xlsx, .xit	<100GB	/

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:

T4.2: An sgRNA library targeting each protein-coding gene as well as ~10,000 intergenic regions of MG1655 (Nat Commun (2018) 9:2475) was purchased from Addgene and is readily available in the lab.

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 tech transfer will be discussed with the KU Leuven Research & Development - Tech Transfer Office. Valorization potential includes licensing of strains or information on linking a specific sequence variant to a phenotype.

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

Existing agreements between VIB and KU Leuven do not restrict publication of data. There is no IP on the generated strains that would prevent us from storing the strains, performing the anticipated experiments or publishing the results.

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.

- Yes

Materials requested from other labs (e.g., reporters) might be subject to MTAs. Implementation of these materials will be done in consultation with our host institution's legal departments to minimize restrictions on their use.

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).

- **Biological material**: Cryotubes and multi-well plates will be labeled with a reference number that links to an entry in our Microsoft Access Database which is hosted on a central server and accessible to all people involved in the project. All relevant information on the specific strains will be included in this database. This includes strain identifier, a clear description of how the mutants were constructed and a link to a whole genome sequence if applicable.
- **Experimental results**: Data will be generated following standardized protocols which are stored in a central OneNote notebook. Furthermore, an E-notebook will be used to register day-by-day activities. Raw data, history and context of experiments, protocols and analyzed data will be uploaded to this E-Notebook and backed up in the cloud. After publication or upon submission of manuscripts for publication, all datasets described in the publication will be deposited in dedicated data repositories (see below).

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.

- Yes

Various data types come with their own metadata containing technical information about settings, machine types, pixel density, resolution, channels... Examples of these include .fastq NGS files containing standard metadata on sequencing technique, or .nd2 following the Nikon metadata standards. Throughout the project, these data files will be preserved with their original metadata. For .txt, .csv, .xlsx files containing tabular information, extra tabs or a head text section will be used to explain the data, the meaning of the columns… For others lacking a formally acknowledged metadata standard, Dublin Core Metadata will be used and a readme file will be saved in the same directory of the datafiles to explain the various data files and give a broad overview of the analyses steps. Moreover, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards that are more specific to our data.

After publication or upon submission of manuscripts for publication, all datasets described in the publication will be deposited in data repositories (see below). Depending on the repository that is used, the metadata standard used by that specific repository will be filled in.

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

Where will the data be stored?

- **Biological material**: Cryotubes and multi-well plates will be stored in -80° freezers with restricted access.
- **Experimental results**: An E-Notebook will be used to collect data. Low-volume data, protocols and analyses will subsequently be stored in secure and internally shared folders on university servers with built-in backup and versioning (SharePoint). Although built-in backup systems are in place, password-protected hard drives equipped with anti-virus programs will be used as backup. A network drive will also be used for large-scale data (e.g., NGS data and microscopy data). A copy of these datasets will be made to desktop PCs with large computational power (or to a computing cluster of our host institution) whenever data analyses will be performed. For final datasets that are part of publications or manuscripts posted on preprint servers, datasets will be deposited in publicly available repositories. Depending on the data type, this could be the SRA depository (for NGS data), KU Leuven's own data repository (RDR), Mendeley Data... and, whenever possible or required, data will also be fully shared via the publisher's website. Scripts and code will be stored (and shared after reaching a finality) via GitHub or the KU Leuven GitLab server (https://gitlab.kuleuven.be/).

How will the data be backed up?

- **Biological material**: A backup of critical strains will be stored in the lab of Kevin Verstrepen (Bio-Incubator, Gaston Geenslaan 1, 3001 Leuven).
- **Experimental results**: Data will be stored on KU Leuven’s central servers with automatic daily back-up and version control procedures, and on password-protected 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

- **Biological material**: Sufficient storage is available.
- **Experimental results**: OneDrive at KU Leuven offers 5TB of data per user. Network storage is purchased on a group level and increased whenever needed. GitHub space is currently free of charge and only requires small volumes. External hard drives are cheap for large volumes and are readily available in the lab.

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

- **Biological material**: Unauthorized people do not have access to the strains.
- **Experimental results**: E-notebooks are password protected and data stored in KU Leuven's secure environments are secured by a two factor authorization and frequently changed passwords. External HDD are password-protected and stored in the safety of the lab.

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

- **Biological material**: -80° freezers are currently present in the host lab (costs are covered by general lab expenses).
- **Experimental results**: The costs for large volume storage are limited and covered partly by this project (500 €/y) and partly by general lab expenses.

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 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)?

- **Biological material**: All strains will be stored for at least 10 more years after the end of the project. For this purpose, -80° freezers are available in the host lab. Relevant strains will also be deposited in a public repository (e.g., the Belgian Coordinated Collections of Micro-organisms (BCCM)).
- **Experimental results**: Data will in first instance be stored on KU Leuven central servers, and after publication, data will additionally indefinitely be stored in open access repositories (e.g., Zenodo, Mendeley Data, KU Leuven's RDR). Dedicated repositories will be used for specific datatypes, e.g., SRA for NGS data.

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

- **Biological material**: -80° freezers are present (included in general lab costs). Deposit of biological material in public repositories is generally without a fee.
- **Experimental results**: The costs will be covered by general lab budgets.

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.

- Other, please specify:

All published data will be made available at the time of publication. However, in case we identify valuable IP, we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use. This will be done

after consulting with KU Leuven LRD.
Unpublished, essential data will be available to (future) lab members via internal IT provisions.

If access is restricted, please specify who will be able to access the data and under what conditions.

All published data will be made available at the time of publication. However, in case we identify valuable IP, we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use. This will be done after consulting with KU Leuven LRD.
Unpublished, essential data will be available to (future) lab members via internal IT provisions.

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.

- No

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

As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific research outputs are detailed below.

- **Manuscripts**: We opt for open access publications where possible. Publications will be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCiD iD.
- **Biological data**: Bacteria will be shared upon simple request following publication unless we identify valuable IP. In this case, we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use.
- **Research documentation**: All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents deposited in lab notebooks are accessible to the PI and the research staff involved in the project, and will be made available upon request.
- **Algorithms and scripts**: As soon as a manuscript is publicly available, algorithms and scripts will be deposited in a GitHub repository.
- **Datasets**: Datasets (including those of CRISPRi screenings) will be deposited in open access repositories.
- **Nucleic acid and protein sequences**: Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes) and the Protein Database (for protein sequences).

When will the data be made available?

As a general rule all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g., by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed – or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.
Unpublished data will be embargoed for public access for another 5 years to allow the research groups to publish research findings.

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

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and License, with a material transfer agreement when applicable).

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 PID will be added upon deposit in a data repository.

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

It is the intention to minimize data management costs by implementing standard procedures, e.g., for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs are partly covered by this project (500 €/y) and the remaining costs will be covered by the laboratory budget. A budget for publication costs has been requested in this project (2 500 €/y).

6. Responsibilities

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

Jan Michiels

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

Jan Michiels

Who will manage data preservation and sharing?

Jan Michiels

Who will update and implement this DMP?

Jan Michiels