

DMP title

Project Name PersiSTRESS: How the stress-resistance antiporter GadC integrates bacterial osmoregulation and pH homeostasis with antibiotic-tolerant persistence (FWO DMP) - DMP title

Project Identifier 88824

Grant Title 1201922N

Principal Investigator / Researcher Bram Van den Bergh

Project Data Contact bram.vandenbergh@kuleuven.be

Institution KU Leuven

1. General Information

Name applicant

Bram Van den Bergh

FWO Project Number & Title

1201922N

PERSISTRESS: HOW THE STRESS-RESISTANCE ANTIPORTER GADC INTEGRATES BACTERIAL OSMOREGULATION AND PH HOMEOSTASIS WITH ANTIBIOTIC-TOLERANT PERSISTENCE

Affiliation

- KU Leuven
- Other

Flanders Institute for Biotechnology - VIB

2. Data description

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

- Generate new data
- Reuse existing 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).

Type of data	format	volume	how created/obtained
<i>WP1 Antibiotic tolerance accross the acid and osmotic space</i>			
Images of plated bacterial cultures	.tiff, .jpeg	10 Gb	created from (antibiotic/stress) survival assays in the lab, and plated using robotic liquid hanlers
Converted bacterial densities	.txt, .csv, .xlsx	1 Gb	created from the images of plated bacterial cultures, converted into bacterial loads using in-house scripts
Growth dynamics accross acid and osmotic stress	.txt, .csv, .xlsx	1 Gb	assessment of growth dynamics based on automatic OD measurements in a micro-titer plate format.
genomic mutants in E. coli	cryo-preseverd bacterial cultures	100	created in our lab and/or requested at other labs.

DNA sequences	.fasta,.ABI	1 Gb	Obtained by Sandersequencing to verify the constructed or requested mutants
images of DNA gels	.tiff	1 Gb	used to verify various steps of mutant construction
DNA oligonucleotides	a few 100 cryopreserved vials and digitally preserved sequences	1 Gb	used to construct DNA material needed for the construction of mutants.
<u>WP2 Genetic network underlying GadC-(in)dependent antibiotic tolerance</u>			
CRISPRi-Seq pools	cryopreserved plasmid stocks and transformed libraries	100	created in our lab/in collaboration and/or requested from other labs and/or acquired from repositories like AddGene.
Next-gen sequencing (NGS) CRISPRi libraries	cryopreserved at -20°C	100-500	created with a two-step PCR in our lab
QC data of NGS libraries	.txt, .csv, .xlsx, .xdrx	10 Gb	created with a combination of nanodrop, qubit and size analyses
Next-gen sequencing data	.fastq (or zipped versions thereof)	2 Tb	millions of short read sequences, per sample, generated by NGS machinery at core facilities (or in-house).
reference sequence data	.fasta	100 Mb	reference sequences obtained from NIH/pubmed and other publicly available databases.
lists of differentially targetted genes and GO enrichments	.txt, .csv, .xlsx	10 Mb	As a result of the analyses of the NGS data, a list of fold changes and p values will be drafted for each sample. Idem ditto for pathway analyses outcomes based on these lists (e.g. GO-enrichments)
<u>WP3 Biochemical characterization of GadC* variants</u>			

membrane protein purification histograms, yield data, stability data,...	.txt, .csv, .xlsx	100 Mb	Obtained in the labs of profs. Slotboom, Poolman and Maglia at the University of Groningen. These tabular data will characterize the purification, quantity, purity and stability of the GadC* protein variants that we will purify
images of protein gels	.tiff	100 Mb	obtained after (SDS) PAGE of purification fractions.
Radio-active accumulation and electric current measurements	.txt, .csv, .xlsx	1 Gb	to assess activity of the GadC* variants, various assays will be executed at the labs of profs. Slotboom, Poolman and Maglia resulting in tabular data
<u>WP4 Physiology of the GadC-dependent mechanism of antibiotic tolerance</u>			
potentiometric traces	.txt, .csv, .xlsx	1 Gb	readouts of ion-selective probes in function of time
HPLC histograms	.txt, .csv, .xlsx	1 Gb	measurement of uptake of glutamate/GABA by the <i>gadC</i> mutants in function of time and acidity. performed in the lab of Prof. De Biase (Sapienza University of Rome).
microscopy images	.nd2, .tiff, .txt.	1 Tb	fluorescence, time-lapse, single-cell microscopy of <i>gadC</i> * mutants maintained in specific (dynamic microfluidic) environments. Further analyses of the data will yield tabular formats of counts and levels of fluorescence/cell shape/...
cytometry data	.fcs, .txt	500 Gb	Single-cell measurements of cell size/shape and fluorescence of markers/dyes/sensors.
<u>WP5 Proof of concept of the generality of GadC-dependent antibiotic tolerance</u>			

genomic mutants of <i>E. coli</i> , <i>Salmonella</i> , <i>Lactobacillus</i> and <i>Brucella</i>	cryopreserved bacterial cultures	50	constructed using CRISPR-cas and other genomic engineering tools in our lab and the labs of Lebeer (UAntwerp) and de Bolle (UNamur)
Images of plated bacterial cultures	.tiff,.jpeg	10 Gb	created from (antibiotic/stress) survival assays in the lab
antibiotic survival data	.txt, .csv, .xlsx	500 Mb	created from the images of plated bacterial cultures, converted into bacterial loads using in-house scripts

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

NA

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

In WP5, I will use an animal model but one based on *Drosophila melanogaster* which is an invertebrate of which the use is not subjected to ethical clearance.

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?

- No

NA

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?

- Yes

Depending on the material requested from other labs, MTAs might need to be signed. This will always be in agreement with KU Leuven/VIB legal department to minimize restrictions on the use of the material/data.

4. Documentation and metadata

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

Most, if not all data files will be named using "yearmonthday_briefTitleOrDescription". All experiments are preformed using an SOP, which is stored/shared in a group OneNote notebook. Furthermore, a digital notebook in OneNote is used to register day-by-day activities in the lab or at the computer.

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 etc. For file types that do not allow such additional annotation data, a readme file will be saved in the same directory of the datafiles to explain all the various data files and give a broad overview of the analyses steps.

Most if not all analyses will be performed in R (or preferably similar open-source software like Python), making use of analysis scripts per type of analyses/research question. These scripts will be annotated with comments and section heads to explain the various steps and outcomes. For final outcomes, I will make use of a result-embedded markdown file, combining code with outcomes.

Various data types (e.g. .fastq, .nd2, .tiff, ...) come with their own metadata containing technical information about settings, machine types, pixel density, resolution, channels,... These data files will be preserved with their original metadata and some of these metadata will be used/accessed within the analyses scripts.

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
- No

For data files that come with their own, standard metadata (.nd2 files follow the Nikon metadata standards, .fastq NGS files contain standard metadata on sequencing technique), I will preserve such metadata. Upon storing some data in publicly available repositories, their metadata will be followed (e.g. SRA database for NGS data).

Depending on the specific experiment type, I will store minimal required information in the text headers or separate tabs, to explain the conditions of the experiment. For example, for the antibiotic sensitivity tests, it will contain the species name, strain name, mutation, antibiotic type and name, concentration, treatment duration, and type of plating assay. For microscopy and cytometry data, finalized data will be preserved along a metadata file describing relevant parameters for the experiment (species and strain name, mutation, time resolution, duration, channels, exposure time, resolution, ...).

5. Data storage and backup during the FWO project

Where will the data be stored?

Data will be generally stored in two or more locations. For small-volume datasets (i.e., tabular data of antibiotic survival), data will be stored on the device where the data is generated (and/or the attached network drives), my personal laptop and a onedrive/sharepoint location. Off-site backups will be frequently made using two external hard drive copies, each on a different location.

A network drive will be used for large-scale data (e.g. NGS data and microscopy data). A copy of these datasets will be made to desktop pc's with large computational power (or to the computing cluster of KU Leuven) 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.

How is backup of the data provided?

The use of Onedrive/sharepoint is backed up via a local copy and a (back-up) copy in the cloud/on the KU leuven network and offers version control to allow easy recovery in case of mistakes. Similarly, network drives of KU Leuven follow daily back-up procedures and offer version control via the built-in features of Windows. In case of more severe data loss or damage, OneDrive allows a swift recovery of data on a separate machine. The IT support at KU Leuven similarly can easily recover data when network drives fail.

For off-site backups of my personal system using external hard drives, I make use of SyncBackFree software once a month which offers a more long-term backup in case of long-term malfunctioning or infections.

Github-stored code and scripts are backed-up in the cloud, and also offers a version control.

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

OneDrive at KU Leuven alone already 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.

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

Two external HDD were bought on my benchfee. Network/cloud based storage is covered by the host lab and organized on a lab/centre level to which each personal contributes a small yearly fee (in my case, through my benchfee).

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

Before reaching a finalized stage, access to data will be secured since data on the network drives are stored in the university's secure environment. OneDrive storage is linked to my personal secured KU Leuven account. Both are secured by a two factor authorization and frequently changed passwords. External HDD are stored in the safety of my home (copy 1) and the lab (copy 2). Upon sharing data (when finalized or during collaborations), academic licenses or DTA/MTA will be put in place which ensure a proper level of security for the use of the data. We will publish open-access provide a free sharing of data at this point to properly educated people.

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, ...).

Data will be stored for 5 years (and longer, complying with the 10 year data preservation rule of KU Leuven) except:

- material that is non-essential to finalized products (e.g. papers): NGS CRISPRiSeq libraries, DNA oligonucleotides, intermediary steps of mutants that were constructed and used.
- publically available data (e.g. DNA sequences to which our data was compared to).

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

For long-term storage, published data will be stored on the network drive and made publically available alongside the publication (if possible) and in a public data repository like Mendeley Data and KU Leuven's RDR. For specific datasets, specialized data repositories will be used (SRA database of NGS data).

Unpublished but essential data (and similarly the essential and published data) will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

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

To store the essential data for 5-10 years, I will make use of publicly and free repositories (published data) and of the KU Leuven long-term, large-volume central storage servers (estimated cost of 100 eur per year).

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

At this moment, no 3rd partie agreements impose a legal restriction of the use/sharing of data. Whenever we request data or collaborate, we aim to keep restriction as low as possible, in

agreement with our legal department.

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

Full datasets, metadata and analyses scripts, will be uploaded alongside the publication whenever possible and uploaded in a data repository (e.g. KU Leuven RDR or others) in all cases. Unpublished, essential data will be available to (future) lab members via internal IT provisions.

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

- In an Open Access repository
- In a restricted access repository

Final, published datasets will be made available (publicly) for reuse via the KU Leuven RDR platform and the SRA repository (for NGS data)

When will the data be made available?

- After an embargo period. Specify the length of the embargo and why this is necessary
- Upon publication of the research results

Unpublished data will be embargoed for public access for another 5 years to allow the research group to publish research findings.

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

Published data will be available to anyone for any purpose, provided that they give appropriate credit to the creators. Unpublished data will be available for everyone with permission, which is everyone with access to the server of the lab (= employee of Centre for Microbiology and Plant Genetics (CMPG))

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

The chosen online repository will be free of charge, network storage at KU Leuven comes with a cost of 100 eur per Tb per year, covered by the host lab after termination of the project.

8. Responsibilities**Who will be responsible for data documentation & metadata?**

Bram Van den Bergh

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

Bram Van den Bergh

Who will be responsible for ensuring data preservation and reuse ?

Bram Van den Bergh

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

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