DMP title

Project Name EOS CRITICAL - DMP title

Project Identifier 40007496

Grant Title G0I1522N

Principal Investigator / Researcher Jan Michiels

Project Data Contact Natalie Verstraeten, Tel +32 16 376643, natalie.verstraeten@kuleuven.be

Description Urinary tract infections (UTIs) are a worldwide health concern and are mainly caused by uropathogenic Escherichia coli (UPEC). Antibiotic therapy failure and the chronic nature of UTIs can be attributed to a small fraction of transiently non-growing, antibiotic-tolerant cells called persisters. An innovative strategy to cure chronic UTIs would be to induce growth resumption in UPEC persisters, thereby re-sensitizing them to conventional antibiotics. However, our current understanding of persister awakening is far from complete, hampering the development of anti-persister drugs. The aim of this project is to study and understand awakening of UPEC persisters in individual cells up to the atomic level. To reach this goal, we have assembled a strong multi-disciplinary consortium with highly complementary expertise. We will identify novel genes and regulatory RNAs involved in awakening by single-cell RNA-seq and by screening a pooled CRISPRi library. Furthermore, we will characterize persister effectors at unprecedented detail by combining genetic, biochemical and structural approaches. Finally, in view of developing therapeutics to stimulate awakening, we will validate our findings in a collection of clinically relevant strains and a model of intracellular infection. Combined, results from this project will lead to ground-breaking new insights in persister awakening and will pave the way for the development of critically needed anti-persister drugs to effectively clear UTIs.

Institution KU Leuven

1. General Information

Name applicant

Jan Michiels - KU Leuven

Wim Versées - Vrije Universiteit Brussel

Steven Ballet - Vrije Universiteit Brussel

Françoise Van Bambeke - Université Catholique de Louvain

Régis Hallez - Université de Namur

Jörg Vogel - University of Würzburg

FWO Project Number & Title

deCiphering bacterial peRsIsTence of Individual Cells down to Atomic Level (CRITICAL)

Application ID: 40007496

Reference: GOI1522N

Affiliation

- KU Leuven
- Vrije Universiteit Brussel
- Other

Collaboration between KU Leuven, Vrije Universiteit Brussel, Université Catholique de Louvain, Université de Namur and University of Würzburg. Note that Jan Michiels and Wim Versées are also affiliated to 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).

WP1					
Type of data	**Format**	**Volume**	**How created**	**Partners involved**	**Reuse/new data**
CRISPRi sgRNA library of *E. coli* MG1655	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Wang *et al.* 2018, *Nat Commun* 9: 2475. Sublibraries are commercially available through Addgene (#113134, #113135, #113136, #113137, #113146).	Jan Michiels	Reuse existing data
CRISPRi sgRNA libraries of *E. coli* UTI89 (WT + Obg/(p)ppGpp mutants)	Glycerol stocks frozen at -80 °C	2 mL cryotubes	We will use custom R scripts for design and filtering and check for potential off-target annotation using Cas-OFFinder. On-target activity will be ranked via genome-wide activity profiling. sgRNAs will be cloned into a backbone vector and transformed by electroporation into the strain of interest carrying an inducible copy of dCas9 (*e.g.* Addgene #44249) reaching at least a 20-fold coverage.	Jan Michiels	Generate new data
Amplicon sequencing results from CRISPRi libraries	.fastq	400 mb per sample (>500 samples)	Cells will be lysed and sgRNAs amplified by PCR using primers that contain index and adapter sequences necessary for sequencing. The amplicons will be purified and sequenced with a custom sequencing protocol. The sgRNA sequences will be recovered from the resulting reads using Trimmomatic and mapped onto the respective genomes using Bowtie2. Counts tables will be generated using custom R scripts and MAGeCK will be used to prioritize sgRNAs, genes and pathways.	Jan Michiels	Generate new data

scRNA-seq data originating from awakening *E. coli* UTI89 persisters (WT + Obg/(p)ppGpp mutants)	.txt	500 gb	Awakening cells will be fixed and isolated using fluorescence-activated flow cytometry (FACS), followed by one-tube cell lysis, cDNA synthesis and amplification. We will use the poly(A)-independent multiple annealing and dC-tailing-based quantitative scRNA-seq (MATQ-seq) protocol, enabling the detection of all transcripts including those of low abundance at sufficient sequencing depth per cell.	Jan Michiels Jörg Vogel	Generate new data
E. coli UTI89 mutants containing promoter reporters, deletions and complementation constructs	Glycerol stocks frozen at -80 °C	2 mL cryotubes	DNA of interest will be expressed in UTI89 using the arabinose-inducible ara_{BAD} promoter (P_{BAD}) of pBAD/His A (Invitrogen) and more controllable expression systems are available (*e.g.* aTc-inducible expression). Knockouts will be created using λ Red recombineering or phage transduction (starting from an available knockout library of *E. coli* BW25113 for common genes).	Jan Michiels	Generate new data
Quantification of persister awakening at population level	.pzfx, .xlsx	1 gb	Bacterial cultures of (knockout or overexpression) mutants and the WT will be treated with antibiotics to kill non-persisters. Surviving persisters will subsequently be washed, resuspended in fresh medium supplemented with antibiotics and quantified at dedicated time points by plate counting. In this setup, the decline of persisters is indicative of persister awakening.	Jan Michiels	Generate new data

Quantification of persister awakening at single-cell level	.nd, .cvs, .xlsx	200 gb	Persister awakening will be evaluated at single-cell level using the “mother machine”, time-resolved microscopy and flow cytometry.	Jan Michiels	Generate new data
WP2					
Type of data	**Format**	**Volume**	**How created**	**Partners involved**	**Reuse/new data**
Obg interaction partners identified by interactomics	.txt	10 kb	In-house performed *in vivo* proteomics-based photo-crosslinking interactomics study.	Jan Michiels Wim Versées	Reuse our own existing data
E. coli strains overexpressing Obg interaction partners	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Encoding genes will be expressed from strong, tightly controlled promoters, *e.g.* pET systems.	Wim Versées	Generate new data
Binding affinities (K_D) and/or binding kinetics (k_{on}, k_{off}) between Obg and its interactors + the influence of the nucleotide-state of Obg on the affinity + interaction with specific domains	.cvs, .txt	500 mb	Determined using one of the various biophysical methods available in our lab (ITC, biolayer interferometry, microscale thermophoresis, fluorescence spectroscopy).	Wim Versées	Generate new data
Binding stoichiometry of Obg with its interactors	.cvs, .txt	500 mb	Established via in-house available state-of-the-art methods such as size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS, Wyatt miniDAWN TREOS II instrument coupled to Waters HPLC) or mass photometry (Refeyn One).	Wim Versées	Generate new data
Structural analysis of Obg-interactor complexes	.jpeg, .tif, h5 files, .mt, .mtz, .mrc, .hdf, .pdb, mmCIF	2 tb	X-ray crystallography or cryo-EM (Megabody technology), supplemented with low resolution structural information from SAXS.	Wim Versées	Generate new data
Influence of the binding of interactors on Obg’s GDP release and GTPase activity	.cvs, .txt	500 mb	Stopped-flow and HPLC assays.	Wim Versées	Generate new data
Effect of Obg binding on the activity of interacting proteins	.cvs, .txt	500 mb	Unknown at this point.	Wim Versées	Generate new data

Mutants of Obg and its interactors, no longer capable of forming the complex	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Site-specific mutagenesis followed by mutant gene expression, either from a plasmid or from the chromosome.	Wim Versées	Generate new data
[Modified/optimized] peptides/peptidomimetics sequences that specifically affect the interaction between Obg and its interactors without disrupting other functions	.xyz, .pdb, .moe, .dcd, .csv	1 tb	The design of these molecules will be supported by in-house *in silico* modelling (incl. molecular dynamics using MOE) and empirical scoring functions to estimate the binding affinity of designed ligands	Steven Ballet	Generate new data
[Modified/optimized] peptides/peptidomimetics that specifically affect the interaction between Obg and its interactors without disrupting other functions	Lyophilized powder 5 to 100 mgs kept in 4°C fridge	Capped 10 ml glass vials	Synthesized using standard solid-phase peptide synthesis (SPPS) protocols and further optimized and stabilized by cyclization/rigidification.	Steven Ballet	Generate new data
Ability of optimized peptides/peptidomimetics to disrupt interactions of Obg and its interaction partners	.cvs, .txt	500 mb	Determined using one of the various biophysical methods available in our lab (ITC, biolayer interferometry, microscale thermophoresis, fluorescence spectroscopy).	Wim Versées	Generate new data
Structures of relevant peptide-target protein complexes	.jpeg, .tif, h5 files, .mt, .mtz, .mrc, .hdf, .pdb, mmCIF	1 tb	X-ray crystallography or cryo-EM (Megabody technology), supplemented with low resolution structural information from SAXS.	Wim Versées	Generate new data
WP3					
Type of data	**Format**	**Volume**	**How created**	**Partners involved**	**Reuse/new data**
Library of potential (p)ppGpp biosensors	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Genetic engineering and domain-insertion profiling.	Régis Hallez	Generate new data
(p)ppGpp levels quantified using an optimized (p)ppGpp biosensor and reporter strains for various processes including DNA replication and cell division.	.xlxs	50 mb	FACS measurements.	Régis Hallez	Generate new data

UTI89 mutants affected in transcription, translation and purine synthesis to identify (p)ppGpp targets	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Mutants will be constructed using established protocols including recombineering and CRISPRi.	Régis Hallez	Generate new data
scRNA-seq data originating from awakening RNAP* mutants (unable to bind (p)ppGpp)	.txt	100 gb	Awakening cells will be fixed and isolated using fluorescence-activated flow cytometry (FACS), followed by one-tube cell lysis, cDNA synthesis and amplification. We will use the poly(A)-independent multiple annealing and dC-tailing-based quantitative scRNA-seq (MATQ-seq) protocol, enabling the detection of all transcripts including those of low abundance at sufficient sequencing depth per cell.	Régis Hallez Jörg Vogel	Generate new data
Quantification of persister awakening at population level of UTI89 mutants affected in transcription, translation and purine synthesis to identify (p)ppGpp targets	.pzfx, .xlsx	1 gb	Bacterial cultures of mutants will be treated with antibiotics to kill non-persisters. Surviving persisters will subsequently be washed, resuspended in fresh medium supplemented with antibiotics and quantified at dedicated time points by plate counting. In this setup, the decline of persisters is indicative of persister awakening.	Régis Hallez	Generate new data
Identification of (p)ppGpp targets	. xlsx	1 gb	An in-house affinity-based (p)ppGpp capture system [unpublished] will be used to fish proteins preferentially bound by (p)ppGpp in persisters. For that purpose, we will enrich persister cells by filtration before capturing (p)ppGpp targets and identifying them by mass spectrometry (LC-MS/MS).	Régis Hallez	Generate new data

New regulators of SpoT identified by amplicon sequencing of CRISPRi libraries (CRISPRi on *relA* and *spoT* mutants, different nutrient statuses)	.fastq	400 mb per sample (>500 samples)	Cells will be lysed and sgRNAs amplified by PCR using primers that contain index and adapter sequences necessary for sequencing. The amplicons will be purified and sequenced with a custom sequencing protocol. The sgRNA sequences will be recovered from the resulting reads using Trimmomatic and mapped onto the respective genomes using Bowtie2. Counts tables will be generated using custom R scripts and MAGeCK will be used to prioritize sgRNAs, genes and pathways.	Régis Hallez	Generate new data
SpoT interaction partners	. xlsx	1 gb	We will use Tandem Affinity Purification (TAP) of SpoT-TAP fusions followed by identification by LC-MS/MS.	Régis Hallez	Generate new data
SpoT mutants identified by genetic and biochemical screens	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Mutants will be constructed using established protocols including recombineering.	Régis Hallez	Generate new data
Quantification of persister awakening at population level	.pzfx, .xlsx	1 gb	Bacterial cultures of SpoT mutants and the WT will be treated with antibiotics to kill non-persisters. Surviving persisters will subsequently be washed, resuspended in fresh medium supplemented with antibiotics and quantified at dedicated time points by plate counting. In this setup, the decline of persisters is indicative of persister awakening.	Régis Hallez	Generate new data
Quantification of persister awakening at single-cell level	.nd, .cvs, .xlsx	200 gb	Persister awakening will also be evaluated at single-cell level using the “mother machine”, time-resolved microscopy and flow cytometry.	Régis Hallez	Generate new data

In vitro hydrolase activity of SpoT	. xlsx	200 mb	*in vitro* hydrolase assays will be done with SpoT in the presence of the potential regulators.	Régis Hallez	Generate new data
WP4					
Type of data	**Format**	**Volume**	**How created**	**Partners involved**	**Reuse/new data**
Characterization of newly identified awakening effectors	Unknown at this point	Unknown at this point	Unknown at this point (will depend on the nature of the specific effectors), likely genetic, biochemical and expression analyses (at population and single-cell level), the identification of interaction partners, and a molecular characterization	Jan Michiels Wim Versées Régis Hallez	Generate new data
WP5					
Type of data	**Format**	**Volume**	**How created**	**Partners involved**	**Reuse/new data**
Collection of UPECs isolated from urine samples	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Urine samples are being collected from patients with suprapubic catheter during visits to the urology department (UZ Leuven) with 6-week intervals (10 samples per patients, 15 patients). Bacterial strains will be isolated and presumptively identified by plating on CHROMagar™ Orientation and further validation by biochemical and immunological testing and 16S rRNA sequencing.	Jan Michiels	Reuse data that are currently being collected in the frame of a parallel project (ethical clearance: S65007)
Collection of 50/50 non-clonal isolates from patients with persistent/cured (control) UTIs in partner hospitals	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Collected in collaboration with AZ Delta Roselaere and Saint Luc.	Françoise Van Bambeke	Generate new data

Quantification of persister awakening at population level for all strains in the strain collections + 5 mutants (overexpression, deletion) based on results from WP1-3	.pzfx, .xlsx	1 gb	Bacterial cultures will be treated with antibiotics to kill non-persisters. Surviving persisters will subsequently be washed, resuspended in fresh medium supplemented with antibiotics and quantified at dedicated time points by plate counting. In this setup, the decline of persisters is indicative of persister awakening.	Françoise Van Bambeke	Generate new data
Quantification of persister awakening at single-cell level for all strains in the strain collections + 5 mutants (overexpression, deletion) based on results from WP1-3	.nd, .cvs, .xlsx	200 gb	Persister awakening will also be evaluated at single-cell level using the “mother machine”, time-resolved microscopy and flow cytometry.	Françoise Van Bambeke	Generate new data
Expression of identified effectors of awakening in max 20 strains from the strain collections	.nd, .cvs, .xlsx	100 gb	Expression will be evaluated at single-cell level using the “mother machine”, time-resolved microscopy and flow cytometry	Françoise Van Bambeke	Generate new data
Tools for bioluminescence resonance energy transfer (BRET) assay	Glycerol stocks frozen at -80 °C	2 mL cryotubes	Target proteins identified *in vitro* will genetically be fused to nanoLuciferase (to be used together with its substrate furimazine), and the peptide-based tool is chemically fused to an acceptor fluorophore (e.g. TAMRA, BODIPY or AlexaFluor), to generate a fluorescent signal when the 2 molecules are in close proximity (10-100 Å).	Wim Versées Steven Ballet	Generate new data
Evaluation of peptide-based tool compounds (identified in WP2) disrupting interactions between proteins involved in persistence/awakening in UTI89 and selected strains from the strain collections.	.nd, .cvs, .xlsx	100 gb	Compounds will be evaluated at single-cell level using time-resolved microscopy and flow cytometry	Françoise Van Bambeke	Generate new data
Intracellular model of infection adapted to *E. coli*	.docx	1 mb	An in-house developed model using monocytes/macrophages will be adapted to uroepithelial cells.	Françoise Van Bambeke	Generate new data

Evolution of intracellular bacterial counts over time and measurement of the division of individual bacteria	.nd, .cvs, .xlsx	5 gb	We will study the kinetics of regrowth of WT and mutant (W1 and 3) UTI89 strains after drug removal, looking at the delay before awakening and the growth rate as well as the possible heterogeneity of the resulting population. In parallel, we will follow the intracellular fate (subcellular localization) and division of bacteria using time-lapse confocal microscopy to evidence possible sub-populations of dividing and non-dividing bacteria, and determine their respective subcellular localisation (cytosolic *vs*. vesicular; with appropriate labelling of acidic vacuoles). Lastly, we will examine the capacity of peptide-based compounds to wake up intracellular bacteria.	Françoise Van Bambeke	Generate new data
scRNA-seqs analysis of intracellular bacteria	.txt	500 gb	Individual bacteria will also be isolated and subjected to scRNA-seq to evidence the activation of specific pathways depending on their state of persistence/awakening.	Françoise Van Bambeke Jörg Vogel	Generate new data
Pharmacodynamic model	.nd, .cvs, .xlsx	5 gb	Develop a pharmacodynamic model (hollow fiber) mimicking antibiotic pharmacokinetics against both extra- and intracellular infections. Evolution of the proportion and dormancy depth of persisters will be followed in this model.	Françoise Van Bambeke	Generate new data

Dual RNA-seq in infected cells	.txt	500 gb	We will perform this analysis in conditions of antibiotic exposure for which the majority of the surviving bacteria are in a dormant state or in the early stages of awakening, in comparison with conditions of active intracellular multiplication.	Françoise Van Bambeke Jörg Vogel	Generate new data

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.

- Yes

This study was registered at CTC with s-number S65007
We are collecting bacterial strains from patients with suprapubic catheters that visit the hospital at regular time intervals (6 weeks). During these visits, urine samples are collected and stored at -80 °C, and patients are presented with a questionnaire inquiring about use of antibiotics, age, sex, and the date of first placement of the catheter. Later, bacterial strains are isolated from the urine samples and phenotypic analyses are performed as described in the project proposal.

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)

- Yes

The collection of urine samples and personal information (i.e. antibiotic use, age, sex, date of placement of catheter) was approved by the Ethics Committee Research UZ/KU Leuven on March 4 2021 (reference number S65007).
Note that in WP5 we propose to use either J774 cells, which are mouse macrophages, or THP-1 human monocytes or T24 ATCC® HTB-4™ cells from human bladder. The two human cell lines are commercial cells lines and no ethical approval is required for use.

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 tech transfer will be discussed with the research and development offices of KU Leuven, Vrije Universiteit Brussel, Université Catholique de Louvain, Université de Namur, University of Würzburg and VIB. Ownership of the generated data has been stipulated
in a Consortium Agreement.

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

Materials requested from other labs (e.g. reporters needed to perform the work described in WP4) might be subject to MTAs. This will be done in consultation with our host institutions' legal departments to minimize resistrictions on the use of these materials.

4. Documentation and metadata

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

BIOLOGICAL MATERIAL
Cryotubes 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 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? 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

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 etc. 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 all 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. Moreover, datasets will need to be registered in VUB’s cris system PURE, where the FOSB (Flemish Open Science Board) metadata schema is used, which is based on the Data Cite metadata standard.

5. Data storage and backup during the FWO project

Where will the data be stored?

BIOLOGICAL MATERIAL
Cryotubes 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 computing cluster of our host institutions) whenever data analyses will be performed. Structural biology data will also be stored at the synchrotron or at the VIB-VUB facility for Bio Electron Cryogenic Microscopy (BECM). 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, GEO, PDB, EMDB... 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.

NOTE ON PERSONAL DATA

Personal data will be pseudonymized using tools supplied by UZ Leuven. Pseudonymised personal data will be stored in a password-protected Excel sheet on a secure KU Leuven server with daily backup. Only investigators involved in the study will be granted access to these files.

How is backup of the data provided?

BIOLOGICAL MATERIAL

A backup of critical strains will be stored in other host labs involved in the project.

EXPERIMENTAL RESULTS

Data will be stored on universities' central servers with automatic daily back-up and version control procedures.

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.

- No

BIOLOGICAL MATERIAL

Sufficient storage is available or will be installed shortly (an additional -80 °C freezer for the Van Bambeke lab is budgeted on the project).

EXPERIMENTAL RESULTS

OneDrive at KU Leuven or SharePoint at VUB 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 labs.

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

BIOLOGICAL MATERIAL

-80° freezers are currently present in the host labs (costs are covered by general lab expenses) or will shortly be installed (Van Bambeke lab, budgeted on the project).

EXPERIMENTAL RESULTS

The costs for large volume storage are covered by the project (€1000 per year).

Data security: 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 the universities’ 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 labs.

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 will be retained at least for 10 years, as required by our host institutions' RDM policy.

Where will the data be archived (= stored for the longer 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 different host labs. 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 the universities' 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. As another example, structural biology datasets will be submitted to the Protein Data Bank (www.wwpdb.org), the Small aAngle Scattering Biological Data Bank (www.sasbdb.org) or the Electron Microscopy Data Bank (www.ebi.ac.uk/pdbe/emdb). The PDB provides both metadata and data according to properties defined in the PDB Exchange Dictionary and the Macromolecular Crystallographic Information Framework (mmCIF). SASBDB is a curated repository of freely accessible and fully searchable SAS experimental data, which are deposited together with the relevant experimental conditions, sample details, instrument characteristic and derived models. SASBDB consents to import and export data using sasCIF, an extension of core Crystallographic Information File for SAS. The EMDB is a public repository for electron cryo-microscopy volume maps and tomograms of macromolecular complexes and subcellular structures. Every EMDB entry has a header file containing metadata (e.g., sample, detector, microscope, image processing) describing the experiment.

What are the expected costs for data preservation during the retention period of 5 years? How will the 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 (€99,55 per TB per year) will be covered by general lab budgets.

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

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 the research and development offices of our host institutions.

Unpublished, essential data will be available to (future) lab members via internal IT provisions.

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

BIOLOGICAL MATERIAL

Microbial strains generated in this project will be made available upon simple request.

EXPERIMENTAL RESULTS

Published data are likely to include CRISPRi libraries and screening results, scRNA-seq results, dedicated mutants, results from phenotypic analyses (awakening, (p)ppGpp levels), structural analysis data, biochemical parameters, peptides/peptidomimetics, intracellular infection models. These datasets will be made available via dedicated public databases (e.g. SRA, PDB, SASBDB, EMDB) or made available via open access repositories (e.g. Zenodo) (see above).

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

- Other (specify):

BIOLOGICAL MATERIAL

We aim at communicating our results in top journals that require full disclosure of all included data. Biological material will be shared upon simple request following publication.

EXPERIMENTAL RESULTS

We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply. Proper links to these data sets will be provided in the corresponding publications.

When will the data be made available?

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?

BIOLOGICAL MATERIAL

Biological material will be distributed to other parties if requested.

EXPERIMENTAL RESULTS

Depending on the journal: open access or standard subscription-based publication.

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

BIOLOGICAL MATERIAL

Generally, shipment is paid by requesting parties.

EXPERIMENTAL RESULTS

A budget for publication costs has been requested in this project (€45000).

8. Responsibilities

Who will be responsible for data documentation & metadata?

All researchers will regularly upload their data, protocols and metadata on the E-notebook and university servers, and regular back-ups will be additionally made on external hard disks (password protected and encrypted).

The final responsibility for data documentation & metadata will be with the WP leaders:

- Jan Michiels: WP1 and WP4
- Wim Versées WP2
- Régis Hallez: WP3
- Françoise Van Bambeke: WP5

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

WP leaders will be responsible for data storage & back up during the project:

- Jan Michiels: WP1 and WP4
- Wim Versées WP2
- Régis Hallez: WP3
- Françoise Van Bambeke: WP5

Who will be responsible for ensuring data preservation and reuse ?

WP leaders will be responsible for data preservation and reuse:

- Jan Michiels: WP1 and WP4
- Wim Versées WP2
- Régis Hallez: WP3
- Françoise Van Bambeke: WP5

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

The PI (Jan Michiels) bears the end responsibility of updating & implementing this DMP.