DMP title

Project Name DMP_11J6222N_Bacterial protein aggregates as novel tools in synbio applications - DMP title

Project Identifier 11J6222N

Principal Investigator / Researcher Abram Aertsen / Ronald Van Eyken

Project Data Contact ronald.vaneyken@kuleuven.be, +32470824538

Description The goal of this project is to employ protein aggregates and their known fundamental principles in E. coli for engineering of novel synthetic behavior in bacterial cells and populations. Therefore, genetic constructs and their specific phenotypic effect on engineered bacterial strains will be closely monitored and analyzed to provide a proof-of-concept for protein aggregates as a novel tool in the synthetic biological toolbox for more complex metabolic engineering and biocontainment strategies. The main types of data collected in this study are: (time-lapse) microscopy images, genetic sequences and the quantification of both bacterial growth and lycopene production.

Institution KU Leuven

1. General Information

Name applicant

Ronald Van Eyken

FWO Project Number & Title

11J6222N - Bacterial protein aggregates as novel tools in synbio applications

Affiliation

- KU Leuven

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
Microscopy images	.nd2/.tif	1-4Tb	Phase-contrast and fluorescence (timelapse) microscopy of bacterial samples
Genetic sequences and sequencing data	.gb/.fasta/.ab1	1-10Gb	Results from Sanger/whole genome sequencing and design of constructs/primers, e.g. using the "Benchling" portal.
Notes, reports, work documents, presentations	.docx/.xlsx/.pptx Lab-journal	<5Gb	Using the Microsoft Office suite, research data will be analyzed and described: excel sheets and calculations, edited images, presentations, summaries and papers based on the obtained research data will be created. All day-to-day research activities will be reported in a non-digital lab-journal.
Multi well plate readings	.xlsx	<5MB	Multiwell plate readers will quantify the development of optical density and fluorescence over time and report respective numerical values in tables.
Lycopene yield data	.xlsx	<5MB	HPLC readings of lycopene yields after harvesting from a bioreactor equipped with engineered bacterial chassis of this project.
Pictures of microbial culture plates and SDS-PAGE/Western blotting gels	.png/.jpeg	<1Gb	In bulk assessment of constructed bacterial strains to verify them before microscopic single cell imaging. Taken with a phone camera. Proteomic analysis is performed using SDS-PAGE/ Western blotting gels. They will be visualised with a gel-doc imaging system.

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

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

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

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?

- No

4. Documentation and metadata

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

1. All microscopy experiments have metadata included into the resulting image file with additional information about the type of experiment, strains, conditions. For time lapse experiments further information about temperature of incubation and specific information of each respective recorded position are gathered, if applicable. Generic metadata about i.e., dimensions, image type, bitdepth, pixel sizes and microscope settings are also included. Experimental specifications i.e., about sample preparations are reported in handwritten notes in lab journals or digitally in the 'metadata' section of every raw microscopy file. Digitalized protocols for individual experimental archetypes are present or will be created for potential later reproduction of results by other individuals.
2. Genetic data is either generated or imported into the "Benchling" portal, a cloud based online platform and respective metadata is added to the files. Additionally all raw genetic sequences are as much as possible annotated to provide detailed information for every genetic construct. Access to data is shared inside the lab, making data internally accessible even after a person has left the lab. Additionally, for all primer sequences and all constructed bacterial strains, a comprehensive list will be provided featuring all the necessary information for further use (annealing temperatures, explanation of the primer's purpose and the primary template for the primer's table; For the strain's table: Species, full genotype, present plasmid(s), growing conditions, comprehensive notes on the construction of the strain, date of construction and location in the -80°C freezer). Sequence data itself is always coupled to a genetic sequence in 'Benchling'. The folder containing the raw sequence data will be dated and given the same name as its cognate genetic sequence.
3. Experimental results and other work documents are readily summarized and/or stored on our institute-based cloud "Onedrive for business", together with text files describing the respective experimental setup and conditions.
4. All day-to-day research activities are without exception completely described in the non-digital lab-journal which is ordered according to date and type of project so that any experiment and its conditions can be quickly found and checked if needed.
5. Any instrument reading (multiplate well reading, HPLC) will be provided with a detailed description in the instruments software or an accompanying document will be written describing the conducted experiment with this instrument in detail.
6. All SDS-PAGE/Western blotting gels and microbial culture plate images will be imported in Powerpoint so that all lanes/bands on gels and all bacterial strains on the culture plates can be annotated properly. These annotations can additionally be deduced from the non-digital lab journal to be found at the date of the experiment mentioned in the name of the Powerpoint file.

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 microscopy data, a metadata model format with fill in fields has been created, so that the blanks can easily be filled in with the respective information. As mentioned before all other generic metadata has a standard format in the NIS-elements software and is provided for every microscopic picture taken.
- Genetic data is not described using a general metadata format. A description of any genetic sequence/construct is provided and annotations of certain parts of the sequence are given. The list of primers, used to create these genetic sequences, and the strainslist summarising strains equipped with these genetic constructs provide additional information.
- A personal semi-metadata system is used to describe day-to-day research activities in the lab journal or in digital summarized experimental results/ work documents: The same type of information is always provided when performing a certain experiment.
  An example:
  - For any PCR that is run, the following information is provided: Type of PCR and additional information if needed for a specific PCR type, Purpose of PCR, Polymerase used, primers used, annealing temperature, template used and length of produced DNA with elongation time of the PCR.
- For instrumental readouts, standard metadata formats have already been established explaining the respective numerical outputs.
- Also for images of SDS-PAGE/Western blotting gels and microbial culture plates, a personal semi-metadata system is used and annotations are made in Powerpoint:
  - SDS-PAGE/ Western blotting: Short description of the experiment, every lane and thus every sample that was run on the gel is annotated with sufficient experimental information, protein ladder with appropriate protein mass, date of experiment, important bands indicated with an arrow and described next of the image, extra information on expected protein mass of samples.
  - Microbial culture plates: Short description of the experiment, involved bacterial strains, dilution factor, date of experiment.

5. Data storage and backup during the FWO project

Where will the data be stored?

All research data will be digitalized (except for lab-journals) or has naturally a digital format.
1. Copies of the research data will be/are stored on a personal computer, 2 external harddrives (each residing in a seperate location: one at the applicant's residence and one at the institute's address) and in the cloud (Onedrive for business or the lab Internal network storage (KULeuven)) in parallel.
2. Microscopy data is saved on the microscopy computer and immediately copied onto a personal computer and at least one of the external hard drives. The large volume server of the lab Internal network storage (KULeuven) will also be used in parallel to store all microscopy data.
3. Genetic sequence data is stored in the cloud (Benchling) for easy shared access.
4. Lab journals are kept in the lab and are stored in an archive once a person has left the lab.

How is backup of the data provided?

Data is backed up periodically via online cloud storage (Onedrive for business).
At regular intervals backups of all data will be created on two large external hard drives (4Tb). This is especially the case for microscopy data which are very large for online clouds. This data will be thus stored at least in twofold to avoid loss of data due to potential failure of one of the harddisks. If capacity permits, microscopy data will also be stored in the lab Internal network storage (KULeuven), certainly for the most crucial microscopy data.
Once projects or chapters of larger projects are completed, i.e. publication of data, said data will be archived in an internal network.

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

Two personal 4Tb hard drives are used to create regular two-fold backups of all research data, especially regular copies of microscopy data.
Onedrive for business cloud services provide 2Tb of storage space and is more than sufficient to store all research data except for all of the microscopy data.
Lab Internal network storage can be expanded if needed but still has sufficient capacity.

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

Ca. 100 Euro/year/person for extended storage on backed-up network drives are allocated. This is typically covered by the budget for computer equipment.

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

None of the data are publicly available and no personal data will be generated in the scope of this project making the data less sensitive.

However, all raw data files will only exist on internal networks or hard drives protected by passwords, respectively, and thus not be accessible by unauthorized persons.

The same goes for all cloud-based data i.e. sequencing data, work documents, etc..

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 important data, namely data from conclusive experiments linked to the eventual dissertation or publications will be stored in an institute-based cloud. Lab journals will physically be stored and kept by the lab manager.
Preliminary data might not be retained in case datasets would be too small to be of statistical relevance and/or in case of easy and low-cost reproducibility.

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

Data and metadata 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.

Lab Journals are physcically archived in the host institution's lab in a secure cabine with an easy access.

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

Ca. 100 Euro/year/person for extended storage on backed-up network drives are allocated. This is typically covered by the budget for computer equipment.

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

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

Experimental data will only be shared inside the lab or with collaborators upon request. In line with the FWO open access obligation, possible publications resulting from this project will be published in open access journals and thus be publicly available. Data related to unpublished work will not be made available and will be internally used for further research.

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

- In a restricted access repository
- Upon request by mail
- Other (specify):

1. The main tool to find and reuse data for the lab is the institute-based cloud service: 'Ondrive for business'. For genetic sequences/constructs and sequencing data specific, the data is/can be shared via the 'Benchling' cloud service.
2. For larger file sizes the internal network will be used.
3. For external collaborators, files can be shared via mail or Onedrive for business.

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

The dissertation resulting from this project and supporting data will typically be held under an embargo for 5 years to protect follow-up projects and related ongoing work.

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

After the embargo has been lifted the dissertation and supporting data will be publicly available. All other research data will remain only accessible to members of the lab unless specific exceptions are granted for non-profit researchers.

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

Minimal costs are expected. These can be covered by budget allocated to miscellaneous costs.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The PhD student pursuing this project is responsible for maintaining sufficient documentation about generated data and creating respective metadata.

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

This obligation is shared between the student and the lab manager. The student creates short term backups, as previously mentioned, and stores the data. Long term storage and backups are being overseen by the lab manager.

Who will be responsible for ensuring data preservation and reuse ?

The PhD student is responsible for initial data preservation and general reusability of data. Once the data has been backed up/archived, the institute´s IT department makes sure the data is preserved and maintains its intended accessibility.

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

The PhD student is responsible for direct implementation & updating of this DMP. The PI bears the end responsibility of this DMP.