DMP title

Project Name 2D-arrays formed by the SAKe designer proteins as a scaffold for catalytic enzymatic cascades - DMP title

Project Identifier u0047937

Grant Title G0C0222N

Principal Investigator / Researcher Arnout Voet

Description In this project we will re-engineer the computaiontally designed SAKe proteins which are able to form 2D layers. We will investigate the formation kinetics and satbility of the layers using biophysical methods including AFM and fluorescence spectroscopy. This information will be used to design varieties with improved stability properties. In parallel, we will functionalize the SAKe proteins with cargo proteins starting from fluorescent proteins to end with enzymatic model systems.

Institution KU Leuven

1. General Information

Name applicant

Arnout Voet

FWO Project Number & Title

G0C0222N
2D-arrays formed by the SAKe designer proteins as a scaffold for catalytic enzymatic cascades

Affiliation

- KU Leuven

2. Data description

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

- Generate new 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
protein scripts	.py	100MB	programming in python and pyrosetta
protein designs models	.pdb	200-300MB	rosetta computational design
expression constructs	sample	20 microliter	custom gene synthesis by biotech company
protein purification data	docx	100MB	experimental reports documenting every step of the process
raw diffraction data	.cbf, h5	1TB	diffraction at synchrotron
protein crystal structures	.pdb,mtz	4-10GB	solving diffraction at synchrotron radiation facilities
AFM images of layers	.ibw	400-600GB	AFM microscopy
biophysical observation of layer formation	.csv	25GB	raw fluorescence measurements
observation of catalysis	.csv	25GB	raw spectroscopic measurements

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

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

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?

- Yes

This deals with a further engineering of a protein for which IP (by the applicant, full freedom to operate) protection is ongoing. The 2D improved 2D layers could be further used in diagnostics, nexzt to potential applications in biocatalysis. Protein sequences may be protected via application pattents. This will be aassisted by LRD.

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. For every protein design the script (.py) and resulting pdb files will be saved and noted down in the digital benchling lab book, with meta data files in the corresponding folders.

2. For every design a protein will be created from a synthetic construct. The genetic code , vector map iwll be stored into the benchling digital lab book. Using Altec label printer and QR barcoding system the sample will be labeled and linked to the digital data indicating storate location.

3. For every protein the full purificaiton details will be written down in the digital lab notebook.

4. Crystallisation of relevant proteins, followed by diffraction and data so lving will be performed. For every protein a fold will be created which contains all crystallographic data and meta data indicating the steps of processing. This progress will also be included into the digital lab book.

5. For every AFM assembly, a folder will be created which contains per experiment a nested folder containing all AFM raw images and analysis. The experimentel details and analysis will be part of meta data but will also be included into the digital lab book.

6.For every biophiyical characterisation of stability (using fluorescence microscopy in flow channels) a separte folder will be created which contains the raw observations (saved as .csv files) with meta data describing the expeirments. The experimental details and analysis will also be included into the digital lab book.

7.For every experiment related to the functionalisation of proteins with enzymatidc cargo protein a new folder will be created which contains the raw observations as .csv files, including meta data to describe the experimental setup and analyasis methodology. The experimental details and analysis will also be included into the digital lab book.

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

In general each lab member will use an electronic lab notebook (benchling) in which a number of predetermined topics have to be described for each experiment (objective, protocol, results, and conclusion). Any lab book format used will contain references to the digital data.
Metadata with the connection between lab samples and files on our data storage so that data files, lab samples, and experimental notes (including descriptions of equipment, setting and used experimental settings) will remain properly linked.
As a general rule, raw data of every experiment type will be sorted and stored in separate folders. Per work package, processed data will be sorted and stored in separated folders with links to their respective raw data files. Separate documents of non-experimental naturewill be sorted and stored in a documents folder.
Any form of publication can require a combination of data from multiple experiments and work packages. Thus, new folders will be made for all processed data per publication (and links to respective raw data).
When depositing data in a local or public repository, the final dataset will be accompanied by all relevant information in a README.txt document, following the Dublin Core Metadata standard if no other meta-standard is available yet. This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of thelaboratory and add contextual value to the dataset for future reuse.

5. Data storage and backup during the FWO project

Where will the data be stored?

Question not answered.

How is backup of the data provided?

In our lab, all important data is stored on a secure access double backup 132Tb NAS (Synology) with cloud based backup.

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

Every project gets a dediczated NAS drive at the start of the project, and costs have been budgetted within the project.

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

they have been budgetted within the awared project.

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

The NAS drive works with secured directories only accessible to the administrator. Otherwise only the researcher working on this project get access. Others have no reading rights.

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, ...).

Question not answered.

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

The data to be retained during 10 years after the project's end are dissemination data (source files of publications, presentations and patents) and the most relevantmeasurement data. A possible exception to this could become the raw imaging data files from the xray diffraction experiments, but only if it will be judged that keeping thereconstructed/processed imaging data is not necessary and the cost of storing the large amount of raw imaging data will be unaffordable.

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

The hardware has been budgetted as a onetime cost, including cloud based backup at a remote location. For he cloud based data preservation, we estimate a cost of 2000 Euro/year. These costs will be covered by funding acquired by the project PIs in the context of other research projects.

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?

Relevant digital data will be published and made available after the end of the project.

Raw data will be made available via online repistories (PDB database, mendeley, github ...)

Data with valuable IP will be protected prior to publication. We will comply with open access regulations of KU Leuven.

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

- In an Open Access repository

1) The source code will be released on GitHub.
2) Strucdtures will be deposited to the DPB database
3) other analysed data will be made available via data repositories. eg mendeley data.
4) other raw data is available upon request

When will the data be made available?

- Upon publication of the research results

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

All data will be made public without restrictions

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

All repositories to be used are free

8. Responsibilities

Who will be responsible for data documentation & metadata?

During the course of the project the Phd students and or postdocs working on the project. The PI will take frequent roles in ensuring correct preservation of the data. Following the end of the project, the PI will be the responsible prerson

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

During the course of the project the Phd students and or postdocs working on the project. The PI will take frequent roles in ensuring correct preservation of the data. Following the end of the project, the PI will be the responsible prerson

Who will be responsible for ensuring data preservation and reuse ?

During the course of the project the Phd students and or postdocs working on the project. The PI will take frequent roles in ensuring correct preservation of the data. Following the end of the project, the PI will be the responsible prerson

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

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