Plan name: C3/21/020 – DMP

ID: Symmetric antibodies for symmetric epitopes: developing and validating the Symbody platform.

Grant number: C3/21/020

Principal Investigator/Researcher: Arnout Voet, Vitor Pinheiro, Debby Thomas

Principal Investigator/Researcher ID: 0000-0002-3329-2703, 0000-0003-2491-0028, 0000-0003-3075-9846

Plan data contact: arnout.voet@kuleuven.be (+32 16 32 44 96)

Description: The global antibody market has been growing at an unprecedented speed. It was valued at USD 130.9 billion in 2020 and is estimated to grow to USD 223.7 billion by the end of 2025, at a CAGR of 11.31%. However, the high costs of monoclonal antibodies and lack of alternative administration routes (i.e. oral) negatively impact patient affordability and quality of life. Researchers created various alternatives such as antibody fragments, Darpins, Alpha bodies and Nanobodies. Recently, we designed an extremely stable and symmetric class of proteins named Symbodies. Symbodies can be decorated with loops of various lengths, retaining good stability at high temperatures or in harsh chemical conditions and are obtained in large quantities from microbes. Furthermore, the symmetric nature of the SAKe proteins allows for circularization of the backbone via a split-intein system, yielding a chemically- and proteolytically-tolerant chassis. Additionally, intrinsic Symbody symmetry might allow for high-avidity symmetry-matched interactions with various therapeutic target molecules such as TNFalpha, VEGF and viral proteins such as the SARS-COV-2 spike protein. Such symmetry-matched interactions yield high affinity and potency via an avidity effect, and might improve target specificity.

The aim of this project is to optimize and validate the symbody platform for development of innovative biological therapies. More specifically, we aim to reach the following research objectives: **(1)** Design of a humanized Symbody core, with optimal pharmacokinetic properties, ready for functionalization, **(2)** Development of a Symbody evolution platform, yielding Symbodies with high affinity for any given symmetric target protein, **(3)** Demonstration of in vivo proof-of-therapeutic concept of a designed anti-TNFalpha Symbody in a mouse colitis model.

1. **General information**
   - **1.1. Name of the project lead (PI)**
     Arnout Voet
   - **1.2. C3 Project number & title**
     C3/21/020 (Symmetric antibodies for symmetric epitopes: developing and validating the Symbody platform.)
2. **Data description**
   - **2.1 Will you generate/collect new data and/or make use of existing data?**
     Generate new data
   - **2.2 What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.**

     WP1. De-immunization/humanization/optimization of the Symbody core scaffold and biophysical Symbody characterization (Voet group)

SOURCE: Protein design and production, biophysical characterization.
TYPE: DNA for microbial protein expression. Purified proteins. Microbial expression host glycerol stocks. Experimental methods and measurements. Scripts.
ESTIMATED VOLUME: 10 TB (approximation)

Types of data	Format	Volume	How created
Protocol and SOPs	.xls .docx .pdf	2 GB	LibreOffice Writer, LibreOffice Calc
Purified proteins	1.5 mL – 2 mL tubes	100 tubes	Protein purification
Glycerol stocks (microbial expression hosts)	1.5 mL tubes	100 tubes	Glycerol stock after cell culture
DNA (plasmids, synthetic constructs)	1.5 mL tubes	100 tubes	In vitro cloning, ordered from companies
Protein structural data (raw xray diffraction images)	.cbf .h5	8 TB (large sizes, but heavily dependent on amount of crystals and file format)	Xray diffraction images obtained at synchrotrons
Protein structural data (processed)	.pdb .fasta .mtz .cif .mmcif .hkl .ASCII .ccp4	20 GB	Coot, Phenix, PyMOL, CCP4i, XDS and other data processing software
PDB files and protein amino acid/DNA sequences (designed models)	.pdb .fasta .txt .pdf	5 GB	MOE, PyMOL, PyRosetta, obabel (bash script), Benchling
Scripts	.sh .py	500 MB	Programmed in bash or Python
Chromatograms (SEC, IEX, HIC, Affinity chromatography, …)	.ASCII .txt .png .jpg	100 GB	Exported from AKTA software. Some are processed in excel, gnuplot or python.
Agarose, SDS, PAGE gel electrophoresis	.tiff .gel and .jpg	5 GB	Gel images
Spectroscopy data	.txt .png .jpg .ASCII .xls	5 GB	Exported from

			relevant equipment software (Tecan, JASCO CD spectrometer, Zetasizer Nano DLS, Nanodrop,…). Some are processed in excel, gnuplot or python.
Bio-Layer Interferometry data	.txt .xls .pdf .jpg .png	1 GB	Exported from relevant equipment software (Sartorius Octet R4/R8). Some are further processed in excel, gnuplot or python.
Crystal images	.png .jpg	5 GB	Photographed with Nikon microscope
Atomic UltraCentrifugation graphs	.xls .png .jpg .txt	1 GB	Exported from relevant equipment software and processed in excel/python/gnuplot

WP2. *In vivo* PK and stability testing of the Symbody core scaffold (Pharmabs)

SOURCE: Polyclonal antibody production & characterization following immunization of rabbits. pharmacokinetic profiling studies of symbody variants and immunogenicity assessment of symbody variants
TYPE: Purified antibodies. Experimental methods and measurements.
ESTIMATED VOLUME: 6 GB

Types of data	**Format**	**Volume**	**How created**
Figures, graphs and processed data	.jpg, .pfzx	5GB	Excel, Graphpad
Protocol and SOPs	.doc(x)	500 Mb	MS Word
Purified antibodies	1 to 10 ml tubes	50 tubes	See above

WP3. Directed evolution of Symbodies (Pinheiro group)

SOURCE: Development, optimization and validation of Symbody display technology
TYPE: DNA constructs (sequence and plasmid), Sanger and NGS data, Biochemical assays (agarose gels, polyacrylamide gels, biophysical assays)
ESTIMATED VOLUME: 3 GB

Types of data	**Format**	**Volume**	**How created**
DNA constructs	.dna .gbk	100 Mb	Snapgene or Benchling
Sanger and NGS data	.seq .ab .fa .fastq	2 GB	Sequencing performed by company
Biochemical assays (agarose gels,	.tiff .gel .jpg .pfzx .ipynb .jl .fcs .xit	2 GB	GE Typhoon scanner, spectrophotometer,

polyacrylamide gels, biophysical assays) and analysis			flow cytometer, GraphPad Prism, Julia, Galaxy

3. Ethical and legal issues

3.1 Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

Yes. We will isolate PBMC from healthy volunteers to setup in vitro immunogenicity tests. All samples and the corresponding clinical data will be encoded according to GDPR guidelines, i.e. the samples and clinical data will be pseudonymized. In that way, personal information of the patients cannot be released to the staff contributing to the project and to any other third parties. Only the concerned clinicians will have this information. A decoding file will be generated (excel) in which the identity of the healthy volunteers is linked to their corresponding number in the study so that identification is possible if needed.

3.2 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. We have already acquired The Ethics Committee Research UZ / KU Leuven (KU Leuven ECD approval P107/2016) approval for preclinical animal experiments evaluating compound bioavailability, half-life, PK profile and dose safety.

3.3 Does your research 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. Our research will generate novel protein compounds with potential applications in diagnostics and therapeutics. We expect that the proposed work will result in research data with the potential for tech transfer and valorization and have already filed priority for a patent protecting the composition of the Symbody family and uses thereof for treatments and diagnosis (EP21183379.3, filed on 02/06/2021). Additionally, new Symbodies with specific biological effects or new methods aiding Symbody development can also be protected. When there is concrete potential for tech transfer, the IP related to these research data will be protected with the support of KU Leuven LRD and the IOF managers supporting this project (dr. Ivo De Baere, dr. Nick Geukens). The IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application, it will be planned so that publications need not be delayed.

3.4 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 regarding reuse and sharing are in place?

There are no 3rd party agreements in place regarding this project.

4. Documentation and metadata

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

For Voet lab:

Each experiment is registered in the (E-) lab journal of the scientist performing the respective experiment. Standard operating procedures (SOPs) have been and will be written for all the techniques used in the lab. Notes will describe the biological/clinical samples used, experimental setup and protocols used, data generated, links to the specific location of sample derivatives and data, as well as the names of the datasets. Data obtained from experiments will be stored in specific folders that also contains a README.txt file explaining the design/protocol, analysis methods, results and labels used in the data analysis file, and a reference to the (E-) lab journal of that particular experiment. The information provided will allow another researcher to follow all steps in the data processing. Also, algorithms, scripts and software usage will be documented. When scripts, algorithms and software tools are finalized, they will be additionally described in manuscripts and/or online git repositories (e.g. GitHub).

For Pinheiro lab:

(Phase I: Assembly of DNA constructs, validation of surface display, library assembly, library testing/display, validation of selection methods. IP development is limited.)

Protocols and experimental data are registered in the cloud-based e-lab notebook used by the group (Benchling). Data is organised in folders by objective with notebooks detailing experimental parameters (description, design, linked resources) and results (input directly or as attachments to the notebook). All raw data (< 10 Mb) is stored on university cloud-based drives with larger datasets maintained in data repositories (e.g. Github) accessible only to members of the Pinheiro group. Periodic backups (yearly) to PDF are carried out and stored on cloud-based university servers (OneDrive).

(Phase II: Isolating binders, IP valuable data is generated)

Only paper based KU Leuven labbooks and in house cloud based lab book systems will be used to log protocols and experimental data. Data organisation is similar.

For Pharmabs:

Protocols, methods and SOPs will be provided and stored as word or pdf documents. The raw data (antibody production & characterization; in vitro & in vivo immunogenicity testing) will be stored as processed test results (excel files). The figures will be generated by Prism (GraphPad) or excel and stored as .jpg and .pfzx or excel files. All activities and protocols are logged in an eLAB journal. All data will be stored on the secured server of KU Leuven in a folder that is only accessbile for the staff of our group working on the project.

4.2 Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Being a highly interdisciplinary project, it is not possible to use a standard metadata system. In general each lab will use an electronic lab notebook / paper based lab book 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 (C321020_WP<nr.>_<description>), processed data will be sorted and stored in separated folders with links to their respective raw data files. Separate documents of non-experimental nature will be sorted and stored in a C321020_documents folder.
- Any form of publication (papers, patents) 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 the laboratory and add contextual value to the dataset for future reuse.

5. Data storage and backup during C3 project

5.1 Where will the data be stored?

Op het einde alles mergen naar de nas

Digital files will be stored temporarily on local computers and hard drives connected to measurement equipment (BLI, AKTA chromatography systems, …). All data relevant to the project will also be stored on:

Short term

- Voet lab: Raw experimental data are stored on both the local machines and backed-up copies are stored on a 64 TB NAS (Synology).
- Pinheiro lab: Raw experimental data are stored in the local machines with back up copy in cloud (KUL OneDrive, shared folder, GitHub or Benchling – depending on intended frequency of access and file size). For this project, they are storing non-sensitive data in Benchling but will move data to the ICTS shared drives for IP sensitive data.
- Pharmabs: Raw experimental data and processed files will be stored on their KU Leuven Onedrive. The decoding key to pseudonymized personal data is always kept at UZ Leuven on their secure servers.

Long term

- All important project related data will be stored in a KU Leuven VSC storage and online git repositories

Physical samples will be stored in fridges and freezers purchased by the PIs and located in their laboratories. All samples will be tracked by the methods prefered by each lab:

- Voet lab: physical samples are stored locally in fridges and freezers located in the laboratory and connected to a campus wide monitoring network. The laboratory uses an Excel format, containing sample name, content description and location in freezer. They are exploring the option to barcode the samples and link them to a registry in Benchling.
- Pinheiro lab: physical samples are stored locally in fridges and freezers located in the laboratory and connected to a campus wide monitoring network. Samples are labelled and barcoded, linking it to a registry in Benchling.
- Pharmabs: physical samples are stored locally in fridges and freezers located in the laboratory and connected to a campus wide monitoring network.

5.2 How will the data be backed up?

In the Voet lab, all important data will be stored on a secure access double backup NAS (Synology). In the other labs, all data belonging to the project is stored on snapshotted, failure-proof systems managed by KUL ICTS.

5.3 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. KU Leuven ICTS and VSC provide sufficient storage and archival capacity during the project. The laboratory of prof. Voet will expand their NAS system with a 32 TB extension. This should be sufficient to cover all data output from this project and its long term (10y) storage.

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

We have estimated a collection of up to 10 TB storage over the duration of the project. We will be using a secure access double backup NAS (Synology) system cover backups and long term storage of data. We estimate storage costs at 1.200 Euro to cover the projects duration and further long term storage (10 y). This cost has been budgeted.

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

The data is secured by the ICTS service of KU Leuven or is securely stored on a Synology NAS system. Confidential data can and will be protected with a password, available only for PI Arnout Voet and Paul Declerck (the latter for personal data). Visitors, MSc thesis students and internship students in the groups as well as other unauthorized persons will not have access to the data on the shared folder. The VSC storage is only accessible to VSC accounts, and the respective volumes will only be accessible to group members. The key to personal/patient information of pseudonymized data will always be kept with Pharmabs on their secured server.

6. **Data preservation after the end of the C3 project**

6.1 Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

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 relevant measurement 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 the reconstructed/processed imaging data is not necessary and the cost of storing the large amount of raw imaging data will be unaffordable.

6.2 Where will these data be archived (= stored for the long term)?

The research data will be stored in a glacial archive storage system after the end of the project. Dissemination data, namely files corresponding to papers and presentations, will be stored on the PCs of involved PIs, and backed-up daily on the departmental server for long term storage. Analysis code will be stored on ICTS-hosted code repositories.

6.3 What are the expected costs for data preservation during these 10 years? How will the costs be covered?

- The volume corresponding to dissemination data is expected to be relatively low, and therefore can be seamlessly embedded in the PIs' allocation on the departmental server.
- For research data, we estimate a total cost of 1.200 EUR for backups and long term storage.

7. Data sharing and re-use

7.1 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 or because of IP potential)?

Data sharing restrictions will apply due to generation of IP. Regular meetings with KU Leuven LRD will be held to evaluate and protect possible IP generated during the project that could lead to valorization actions. If deemed necessary, data that fall under IP will either not be shared, put under embargo, or a suitable licence will be applied to the data when published (e.g. Creative Commons Licence).

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

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

The approach to share data is upon request by e-mail. Due to the data volume, access will then be granted to a restricted access repository.

7.4 When will the data be made available?

As soon as the research results have been published, the data can be made available to other researchers.

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

All project collaborators will be authorized to have access to all obtained digital and physical data after the project. In case the question originates by researchers outside the consortium, the data can be made available upon e-mail request, and on condition that the users agree to give proper credit, such as co-authorship on their papers building on these data. Usage for commercial purposes will require obtaining a license, or equivalent
arrangement.

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

A restricted access repository can be implemented on a free tool, such as Dropbox, up to a certain volume. If this volume does not suffice, time-limited storage will be considered, thus limited to the time needed to download the data. Furthermore, the laboratory of prof. Voet stores essential data in a 64 TB NAS drive with secure cloud-based backup and online access via Synology. Later, all datasets will be made public via relevant databases (eg. PDB, GitHub repositories, Mendeley data repositories).

8. Responsibilities

8.1 Who will be responsible for the data documentation & metadata?

Research and technical staff working on this KUL C3 project will be responsible for the data collection, documentation and metadata. They will be trained in data management. Supervisors will manage the data storage facilities.

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

The PIs, research and technical staff working on this KUL C3 project will be responsible to store the data on the appropriate accommodation provided by KU Leuven. The ICTS service of KU Leuven is responsible for the back-up of the network drives at KU Leuven. The folders will be managed by the supervisors.

8.3 Who will be responsible for ensuring data preservation and sharing?

The PIs of this project will be responsible for the data preservation and eventual reuse of obtained data.

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

The project leader (prof. Arnout Voet) bears the end responsibility of updating and implementing the DMP.