# Screening the morphological landscape of amyloid aggregates with atomic force microscopy and infrared spectroscopy

DATA MANAGEMENT PLAN

# Administrative details

| | |
|---|---|
| **Project Name** | Screening the morphological landscape of amyloid aggregates with atomic force microscopy and infrared spectroscopy |
| **Project Identifier** | 1128822N |
| **Grant Title** | PhD Fellowship fundamental research |
| **Principal Investigators** | Frederic Rousseau, Joost Schymkowitz |
| **Project Data Contact** | Béla Z Schmidt |
| **Project Description** | Amyloids exhibit structural diversity based on their aggregation states (oligomers, seeds, end state fibrils, etc.) and can form distinct structural polymorphs. Crucially, this diversity correlates with different pathological outcomes and rates of disease progression. This project aims at deepening our understanding of the inconsistencies between in vivo and in vitro amyloid formation and how amyloid polymorphism correlates with variable disease phenotypes.<br>Cryo-electron microscopy has been pivotal in determining the structures of amyloid aggregates at atomic resolution but we need a screening method that can facilitate reliable and efficient high-throughput analysis of amyloid polymorphs without extracting the deposits from the tissue.<br>I will focus on the structural polymorphism of two model amyloid- forming proteins (Amyloid-β and Tau) using samples of increasing complexity. Using a state-of-the-art approach (atomic force microscopy-infrared spectroscopy), I will study the impact of reaction conditions during aggregate formation on the structural diversity of amyloid aggregates produced in vitro. These results will be cross-correlated with in situ and ex vivo aggregates obtained from reporter cell lines and animal models. Following an unsupervised computational approach for clustering, I will explore the impact of experimental conditions during fibril formation and extraction on fibril structure, seeding capacity, and the limitations of structural compatibility. |
| **Institution** | KU Leuven |

# 1. General information

| | |
|---|---|
| **Name Applicant** | Wouter Duverger |
| **FWO Project Number** | 1128822N |
| **Project Title (EN)** | Screening the morphological landscape of amyloid aggregates with atomic force microscopy and infrared spectroscopy |
| **Project Title (NL)** | In kaart brengen van het morfologische landschap van amyloïdeaggregaten met atoomkrachtmicroscopie en infraroodspectroscopie |
| **Affiliation** | KU Leuven |

# 2. Data description

**Will you generate/collect new data and/or make use of existing data?**
I will generate new data and make limited use of existing data.

**Describe the origin, type and format of the data (per dataset) and its (estimated) volume, ideally per objective or WP of the project. You might consider using the table in the guidance.**
Refer to the table below.

**Table 1:** Overview of datasets generated in the context of this project.

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **1a** | Protocols for AFM-IR on recombinant samples | Establish method in the host lab | New | Derived_and_compiled_data | Research_documentation | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | MB |
| **1a** | AFM-IR data of recombinant samples | Establish method in the host lab | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **1b** | AFM-IR data analysis scripts | Establish method in the host lab | New | Derived_and_compiled_data | Research_documentation | Python scripts (.py), ImageJ scripts (.ijm), and other plain text files | 100 | MB |
| **1c** | Bulk FTIR data of recombinant samples | Comparison of AFM-IR to bulk FTIR | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 500 | MB |
| **1c** | DLS of recombinant samples | Comparison of AFM-IR to DLS | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 500 | MB |

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **1c** | TEM of recombinant samples | Comparison of AFM-IR to TEM | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **2a** | Recombinant protein aggregated in various conditions | Study the dependence of in vitro aggregation on environmental conditions | New | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 100 | vials |
| **2a** | AFM-IR of recombinant protein aggregated in various conditions | Study the dependence of in vitro aggregation on environmental conditions | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **2b** | AFM-IR of recombinant protein in various stages of aggregation | Study the conformations proteins/peptides go through during aggregations | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **2c** | AFM-IR of recombinant protein aggregated in the presence of biological macromolecules | Study the effect of biological macromolecules on protein aggregation | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **2d** | Different isoforms and post-translational modifications (PTMs) of recombinant protein | Study the effect of isoform and PTMs on recombinant protein | New | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 50 | vials |
| **2d** | AFM-IR of different isoforms and post-translational modifications (PTMs) of recombinant protein | Study the effect of isoform and PTMs on recombinant protein | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **3a** | Cell lines | Characterisation of amyloid aggregates in cells | Existing data | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 10 | vials |
| **3a** | Protocols for AFM-IR on recombinant samples | Optimise method for cell samples | New | Derived_and_compiled_data | Research_documentation | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | MB |

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **3a** | AFM-IR data of cell samples | Characterisation of amyloid aggregates in cells | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **3b** | Animal tissue sections | Characterisation of amyloid aggregates in animal tissues, preparation for human samples | New | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 50 | samples |
| **3b** | Protocols for AFM-IR on animal tissue sections | Optimise method for animal tissue sections, preparation for human samples | New | Derived_and_compiled_data | Research_documentation | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | MB |
| **3b** | AFM-IR data of animal tissue sections | Characterisation of amyloid aggregates in animal tissues, preparation for human samples | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **3b** | Human tissue sections | Characterisation of amyloid aggregates in human health and disease | Existing | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 50 | samples |

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **3b** | Protocols for AFM-IR on human tissue sections | Characterisation of amyloid aggregates in human health and disease | New | Derived_and_compiled_data | Research_documentation | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | MB |
| **3b** | AFM-IR data of human tissue sections | Characterisation of amyloid aggregates in human health and disease | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **3c** | Amyloids from tissue sections amplified by RT-Quic and PMCA | Probing the conformational fidelity of in vitro templating methods | New | Experimental_data | Samples | Biological and chemical samples: live animals, frozen samples in cryovials, samples stored at 4°C. | 50 | vials |
| **3c** | AFM-IR of amyloids from tissue sections amplified by RT-Quic and PMCA | Probing the conformational fidelity of in vitro templating methods | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |
| **3d** | AFM-IR of recombinant protein aggregated in various conditions | Mimicking ex vivo aggregates by optimising in vitro conditions | New | Experimental_data | Spectroscopy_data | Instrument output files in proprietary format (.irb, .axz), comma-separated values files (.csv), electronic image files (.png) | 5 | GB |

| WP | Dataset | Purpose | New or Existing (source) | Data type | Data subtype | Data format | Size | Unit |
|---|---|---|---|---|---|---|---|---|
| **3d** | Optimised protocols to mimick ex vivo aggregates with recombinant material | Mimicking ex vivo aggregates by optimising in vitro conditions | New | Derived_and_compiled_data | Research_documentation | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | MB |
| | Manuscript | Dissemination of results | New | Derived_and_compiled_data | Manuscripts | Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format; | 10 | GB |

## 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 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.**
No. All samples of human origin are anonymized before they come into our possession. No personally identifiable data will be handled in the context of this project.

**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.

Protocols regarding human samples have been approved by the UZ/KU Leuven Ethics Committee Research (EC Research) with reference number S63759.

Animal experiments have been approved by the Ethical Committee for Animal Experimentation (ECD) with reference number P184-2020.

**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.

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorisation. VIB and KU Leuven have a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such 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. Further research beyond the scope of this project may be necessary for developing a strong IP portfolio.

**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?**
Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below). The data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database, and published along with the results.

**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.**
For instrument-specific datasets, additional metadata will be associated with the data file as appropriate.

For other datasets, the metadata will include the following elements:

- **Title:** free text
- **Creator:** Last name, first name, organization
- **Date** and time reference
- **Description:** Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
- **Format:** Details of the file format
- **Identifier:** DOI (when applicable)
- **Access rights:** closed access, embargoed access, restricted access, open access.

The final dataset will be accompanied by a README.txt document. 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 back up during the FWO project

**Where will the data be stored?**

Digital files will be stored either on KU Leuven servers or in shared laboratory folders of an off-site online backup service. The researchers working on the project will have copies of the data files as well as of the derived and compiled data stored on their personal computers.

The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, version control and roll-back capability, which will be used for storage during the project and after. There is a secondary on-campus physical backup of the online storage which synchronizes with the online content with a one-day delay.

Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).

Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

**How is back up of the data provided?**

The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, version control and roll-back capability, which will be used for storage during the project and after. There is a secondary on-campus physical backup of the online storage which synchronizes with the online content with a one-day delay.

**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. The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, which will be used for storage during the project and after.

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

Data storage and backup costs are included in general lab costs. The Switch Lab has a yearly subscription to an off-site online backup service paid from the general budget of the laboratory. The yearly cost of the service is €5.500. This cost includes unlimited data storage, not only the data belonging to the present project.

Electricity costs for the -80°C and -20°C freezers and refrigerators present in the labs as well as the cost of liquid nitrogen cryostorage are included in general lab costs.

**Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**
All notebooks and physical data are stored in the labs. Entry to the lab requires ID-card and key. Access to the digital data is u-number and password controlled.

## 6. Data preservation after 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, ...).**
The minimum preservation term of 5 years after the end of the project will be applied to all datasets.

**Where will the data be archived (= stored for the longer term)?**
Datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication. For all other datasets, long term storage will be ensured as follows:

- Digital datasets will be stored on storage space of an online data-backup service.
- Vectors: Generally, at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C).
- Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

**What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?**
Electricity costs for the -80°C and -20°C freezers and refrigerators present in the labs as well as for in liquid nitrogen cryostorage are included in general lab costs. The cost of the laboratory's professional subscription to the online data backup service is €5.500 per year (€27.500 for 5 years). This cost includes unlimited data storage, not only the data belonging to the present project. Data storage and backup costs are included in general lab costs.

## 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?**
Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. All research outputs supporting publications will be made openly accessible. Depending on their nature, some data may be made available prior to publication, either on an individual basis to interested researchers and/or potential new collaborators, or publicly via repositories (e.g. negative data). 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. Physical data (e.g. cell lines) will be distributed to other parties if requested.

**Where/how will the data be made available for reuse?**
The data will be shared upon request by mail.

Possible ways of sharing the generated data:

- nucleic acid sequences: GenBank (https://www.ncbi.nlm.nih.gov/genbank/)
- protein sequences: UniProt KB (https://www.uniprot.org/)
- vectors: AddGene (http://www.addgene.org/depositing/start-deposit/)
- cell lines: direct mailing on dry ice
- microscope images: Image Data Resource (http://idr.openmicroscopy.org/about/)
- proteomics data: PRIDE (https://www.ebi.ac.uk/pride/)
- manuscripts: bioRxiv (https://www.biorxiv.org/)
- other digital data: Zenodo data repository (https://zenodo.org/)

**When will the data be made available?**
Upon publication of the research results.

Generally, research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements–note that patent application filing will be planned so that publications need not be delayed–or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

**Who will be able to access the data and under what conditions?**
Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

**What are the expected costs for data sharing? How will the costs be covered?**
It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be covered by the laboratory budget.
The receiving party will pay for sharing physical data (e.g. cell lines).

## 8. Responsibilities

**Who will be responsible for data documentation & metadata?**
Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets.

**Who will be responsible for data storage & back up during the project?**
The research and technical staff will ensure data storage and back up, with support from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.

**Who will be responsible for ensuring data preservation and reuse ?**
The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.

**Who bears the end responsibility for updating & implementing this DMP?**
The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.