Plan Overview

A Data Management Plan created using DMPonline.be

Title: Open-source robotics-assisted synthesis of libraries of SF-bond containing compounds for medicinal chemistry applications

Creator: Wim De Borggraeve

Principal Investigator: Wim De Borggraeve, n.n., Daniëlle Copmans

Data Manager: Wim De Borggraeve, n.n., Daniëlle Copmans

Project Administrator: Wim De Borggraeve, n.n., Daniëlle Copmans, n.n.

Affiliation: KU Leuven (KUL)

Funder: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Wim De Borggraeve https://orcid.org/0000-0001-7813-6192, n.n. n.n., Daniëlle Copmans

Data Manager: Wim De Borggraeve https://orcid.org/0000-0001-7813-6192, n.n. n.n., Daniëlle Copmans

Project abstract:

In this project, we are initiating the development of a semi-automated synthesis platform utilizing open-source robotics components. The platform's primary purpose is to synthesize novel bioisosteres with S-F bonds inspired by recently identified compounds exhibiting antiepileptic properties. The goal is to enhance the biological properties of these compounds. Additionally, we will create new libraries comprising S-F bond-containing compounds for biological evaluation as potential anti-epileptic agents. These compounds will undergo testing through target-based screening and in an animal model of epilepsy.

ID: 209212

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 20-12-2024

Grant number / URL: C2E/24/012

Open-source robotics-assisted synthesis of libraries of SF-bond containing compounds for medicinal chemistry applications

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

TYPE	FORMAT	VOLUME ESTIMATE	HOW CREATED	REMARKS
CHEMISTRY EXPERIMENT DATA	Information about chemical reactions carried out during the course of the project. Specifications on amounts of products used, conversion, yields, purity, temperature, observations, ...		The data will be written down in a paper notebook and transferred to the Electronic Lab Notebook MBook. The ELN will also be the link between the experiment and the characterisation data for the compounds synthesized.	commercial platform for lab note storage
NMR	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition and processing parameters. **Processed data** is in the form of PDF files, images or textual interpretation of the spectra ready for implementation in a paper (e.g. 1H NMR (D2O, 300 MHz): δ 1.61 (s, 6H), 1.68 (s, 3H), 1.72 (s, 3H), 2.17-1.99 (m, 8H), 4.45 (d ofd,2H,JH,H =6Hz,JP,H =6Hz),5.23-5.15(m,2H),5.46(t,1H,J=6Hz).)	less than 50 MB per compound on average (1D spectra 1 MB, 2D spectra 10-40 MB)	Spectra are recorded on the NMR spectrometers (https://nmrfacilities.chem.kuleuven.be/instrument_specifications.php)	Typically less than 5 GB per PhD thesis.
IR	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition and processing parameters as .dpt. **Processed data** is in the form of PDF files, images or textual interpretation of the spectra ready for implementation in a paper	less than 100 KB per compound on average, PDF less than 100 KB per compound	Spectra are recorded on the IR spectrometer	
UV-VIS-NIR	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition parameters, stored as .DSW files. Processed data is available in .xls, PDF or image format.	1 MB per compound	Spectra are recorded on the UV-VIS-NIR spectrometer	
HRMS	Data folders generated by the spectrometer containing **raw** data (usually in proprietary format) and also all relevant acquisition and processing parameters are kept by the colleague recording the data. PDF files with processed spectra.	RAW data typically less than 5 MB per experiment PDF 25 KB per spectrum	Spectra are recorded on the HRMS spectrometer	Typically less than 10 MB per PhD thesis for the PDF files
MS (+ related CHROMATOGRAPHY)	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition and processing parameters. PDF or image files of the **processed chromatogram** is stored in the (electronic) lab notebook for direct implementation in a paper.	Raw data typically is less than 15 MB per experiment	Raw datasets are recorded and spectra are recorded on the respective MS spectrometer (LCMS, GCMS, ESI/MS etc.)	
X-RAY DATA (SINGLE CRYSTAL)	cif, png	1 MB per spectrum for the CIF files	The raw datasets are recorded on the X ray diffractometer. After integration of the reflections and processing, the cif files are provided to the researchers and uploaded to https://www.ccdc.cam.ac.uk in the framework publications (Cambridge Structural Database or Inorganic Crystal Structure Database).	The cif format is a standard format drafted by the international union of pure and applied crystallography. It contains the reflections (indices h,k,l) and observed structure factors and standard deviation.
DLS	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition and processing parameters. **Processed** and **analyzed** data is kept using excel (.xlsx) or prism software (.pzfx)	Raw data exported as .txt file is typically less than 1 Mb per experiment	The raw datasets are recorded on a zetasizer (DLS) apparatus.	
Circular Dichroism	Data folders generated by the spectrometer containing **raw** data and also all relevant acquisition parameters, stored as .jws files which can be converted to .txt files. **Processed** data is available in .xls format.	Raw data exported as .txt is typically less than 1 Mb per experiment	The raw datasets are recorded on a CD spectrometer.	

ELEMENTAL ANALYSIS	Data folders generated by the combustion analyzer containing **raw** data and also all relevant acquisition parameters, stored as .dat files. **Processed** data is avaliable in .txt or pdf format.	Raw data is typically less than 1 Mb per experiment.	Elemental analysis data is generated by a CHNX combustion analyser	Raw data is transferable to ASCII format
THERMAL ANALYSIS	Data folders generated by the TGA or DSC apparatus containing **raw** data (.00X format) and also all relevant acquisition and processing parameters are kept by the colleague recording the data. The **processed data** can be exported to pfd or .jpeg files.	1 MB per cycle. Typically one cycle is required per compound.	Thermograms are recorded on either a thermogravimetric analyzer or a DSC apparatus.	An archive drive is used to store the raw data.
CHEMICAL COMPOUNDS	Vials containing a few mg of compounds. Not for all compounds, mainly final products.		chemical synthesis	stored in Leuven Chem&Tech in stock room 01.186. Inventory in cheminventory cloud software
BIOLOGICAL EVALUATION DATA	Information on biological activity. Various formats: xls, datawarrior. Relevant processed data kept by researcher. Source data kept by the collaborators generating the data.			
(WP2a) Calcium flux assay	Measuring intracellular calcium concentration (fluorescence recordings)			
(WP2b) Observational toxicity in zebrafish larva	New digital data: numerical .xls (excel files) and graphical (Graphpad prism files)	<1 GB	Daily monitoring of zebrafish larva in time (maximum tolerated doses and concentrations)	
(WP2b) Behavioral analysis (zebrafish larvae)	New digital data: numerical .xls and video's (by Noldus software)	10 GB	Locomotor recordings (Noldus)	
(WP2b) Electrophysiological seizure analysis (zebrafish larvae)	New digital data (software): abf-files	<100 GB	Local field potential recordings (LFP) - equivalent to electroencephalograms	
(WP2b) Observational toxicity (mice)	New digital data: numerical .xls (excell files) and graphical (Graphpad prism files)	<1 GB	daily monitoring of mice in time (maximum tolerated doses after PI injection)	
(WP2b) Behavioral analysis (mice)	New digital data: numerical .xls and video's (by software)	10 GB	Locomotor recordings	
(WP2b) Electrophysiological seizure analysis (mice)	New digital data: visual (video's by software)	<100 GB	Video EEG recordings	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

We reuse data generated within a fee for service agreement by the CRO Hybrigenics as part of our in-silico investigations. The data is referenced through internal project documentation in .pdf format.
We reuse data generated by ourselves in previous collaborations leading up to this project application. The IP is with ourselves.

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, animal data (Provide ECD reference number below)

We have ethical permission from KU Leuven to breed zebrafish (000/(GS1/GS2)) and to conduct
experiment on mice (ECD P-numbers 034/2024 and 097/2024). When applicable, additional ECD numbers will be requested for.
Of note, for the work on human-derived cell lines (HEK cells) the following S-number is provided by the Ethics Committee Research (EC Research) UZ/KU Leuven: S69621.

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

Chemical structures with biological activity can be subject of patent applications. The discovery of novel antiseizure compounds will strengthen and expand our IP position.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

All chemical synthesis and related experiments will be documented in the Electronic Lab Notebook and will be linked to the reaction or the compound to which they relate. The metadata of the separate experimental datasets will be stored in the DMP datafolder together with the raw/processed data.
Metadata (including trial number, used hardware to collect the data, hardware settings, experiment date, user name, type of experiment, details of equipment, measurement units, etc.) will be added to the datasets for easy identification and data reuse. A data description document will be created to record the data acquisition process and also on how to facilitate interpretation of metadata.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

The metadata describing how experiments were run are generated by the software of the experimental setups.
This information can be retrieved using the acquisition software and typically is in proprietary format except in a few cases where the metadata is in the form of plain text files.
Each experiment will be described in the ELN with reference to the raw/processed data in the DMP project folder.
For the biology-related experiments, please refer to the answer above.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Large Volume Storage
- Other (specify below)

The active data used during the duration of the project will be stored by the researchers in Microsoft OneDrive (250 GB) and in the MBook ELN. The data will be stored in a specified folder structure. Handwritten notes will be collected in a hardcopy lab notebook. At the end of the project, these written notes will be scanned and uploaded to the MBook ELN.
Data that has to be actively used by KU Leuven or external collaborators will also be shared via Microsoft OneDrive.
All project data will be shared with the PI via Microsoft OneDrive and via the ELN.
Analog lab notebooks will be stored by the researcher during the project and handed over to the PI for storage at the end of the project, after having made an electronic pdf copy for storage in the MBook ELN.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

For the active data, the data is automatically backed up via OneDrive.
ELN data will be stored in MBook cloud which provides backup.
At the end of the project, the OneDrive DMP datafolder will be migrated to an Archive storage hosted on KU Leuven servers. This archive storage is backed up automatically.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

Microsoft OneDrive provides 250 GB of storage per user.
Archival storage is rented at the KU Leuven ICTS datacenter and is expanded in function of the needs.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Both OneDrive data and Archival storage at KU Leuven is not publicly accessible and only people with permission (PI and/or his delegate responsible for data storage).
Data generated is stored in lab books and on local computers which are kept in the laboratory and only accessible by approved personnel with a badge. Backup files are kept on University servers. Access to these files can only be granted by the PI of the research group.

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

ELN costs amount to about 150 Euro/user/year
Data storage is rented per terabyte at KU Leuven at about 100 Euro/TB/year

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Other (specify below)

archival storage (K-drive) and LVS
redundant copy on NAS maintained by the PI

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

Data storage is rented per terabyte at KU Leuven at about 100 Euro/TB/year. Costs for data storage are described in the budget.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Other (specify below)

As the data is patentable, it will be restricted until IP is filed.
All processed experimental data needed to write the papers is published in the supporting information of the paper. This is typically also a requirement to publish in high-impact journals.
Raw data is kept available by the PIs involved in the project and can be made available to externals upon request and upon approval of the team of PIs.

If access is restricted, please specify who will be able to access the data and under what conditions.

Only the project PIs and the researchers involved in the research project will be able to access the data. Disclosure is not allowed until patent is filed.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- Yes, intellectual property rights

No data related to compounds with biological activity can be made public until a decision of patent filing has been made. Patent exploration foreseen with LRD in close collaboration with IOF managers and PIs involved.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)
- Other (specify below)

After Open Access publication, corresponding datasets data will be shared on open platforms (e.g. KU Leuven Research Data Repository (RDR), NCBI-SRA, and the Open Science Framework (OSF)). We will explore the possibilities via online repositories and will use the website www.re3data.org.
The data and insights will be, later, made available via publications or patents, which are accessible via LIMO (KU Leuven) and search engines like patentscope.
More detailed information and data can be shared upon request by mail and approval by
responsible PI (main data owner).

When will the data be made available?

- Upon publication of research results
- Other (specify below)

After patent filing.

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

After patent filing, publicly available patent online.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

Data is linked to chemical compounds and separate batches of compounds and does not stand in itself. It does not make sense to make such datasets for each and every compound that will be synthesized.

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

This overlaps with the costs for data storage since in case data needs to be shared, it will be looked up in the archive by the PI and shared after making a local copy via OneDrive or sent via Belnet filesender.

Responsibilities

Who will manage data documentation and metadata during the research project?

The PI/Postdoc/PhD/MSc/Internship researchers involved in the project.

Who will manage data storage and backup during the research project?

The PI/Postdoc/PhD/MSc/Internship researchers involved in the project.

Who will manage data preservation and sharing?

The project promoters will make sure the data is stored on KU Leuven archival storage.

Who will update and implement this DMP?

The project promoters will update and follow up on implementation.