Thiazyl fluorides: introducing S≡N triple bonds into SuFEx warheads for protein profiling

A Data Management Plan created using DMPonline.be

Creator: Joachim Demaerel https://orcid.org/0000-0003-3685-7535

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 12B5024N

ID: 204531

Start date: 01-11-2023

End date: 31-10-2026

Project abstract:

Sulfur(VI)-Fluoride Exchange (SuFEx) chemistry describes the array of highly efficient and selective substitution reactions at S(VI)-F centers. These mild electrophiles are attractive as irreversible warheads in chemical biology, notably in activity-based protein profiling (ABPP) where chemical probes are used to identify and study enzymes via ligation to nucleophilic side chains. Where previous work almost exclusively focused on sulfonyl (SO2) based sulfur(VI)-fluorides, no attention since the rise of SuFEx chemistry has been spent on thiazyl fluorides, S≡N-triple bonded, stable structures with very few known applications. The most promising progenitor to thiazyl derivatives is the parent thiazyl trifluoride (N≡SF3), a highly stable gas. However, no useable lab-scale preparation exists to this hub, and therefore entire classes of azasulfur derivatives are unattainable for lack of synthetic entry. In this project, I will tackle this availability hurdle by developing a universal route to all mono- and difunctionalized thiazyl fluoride derivatives, by using only commercial, solid reagents. Next, I will extend this protocol to include sulfondiimidoyl fluorides, another class of azasulfur structures that have no precedents for substituents other than carbon. Lastly, I will make alkynylated activity-based probes to explore the reactivity of these new warheads towards whole proteomes using gel-based fluorescence readout, and analyze hits using an LC-MS/MS proteomics workflow.

Last modified: 30-04-2024

Thiazyl fluorides: introducing S≡N triple bonds into SuFEx warheads for protein profiling
FWO DMP (Flemish Standard DMP)

1. 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
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 noted in the Electronic Lab Notebook MBook. The ELN will also be the link between the experiment and the characterisation data for the compounds synthesized.	Generate new data	Digital	Experimental, observational	NA		
NMR DATA	Spectral data from 1H, 13C and 19F NMR spectroscopy. Data folders are 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.	Generate new data	Digital	Experimental	.pdf, .jpg, .ppt	<100GB	

IR DATA	Spectral data from infrared spectroscopy. Data folders are generated by the spectrometer software 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	Generate new data	Digital	Experimental	.dpt; .jpg, .pdf	<1GB	
HRMS DATA	Data generated during submitted measurements at the high-resolution mass spectrometry facilities. 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 operator recording the data. PDF files with **processed** spectra are sent to the submitter and attached to the electronic lab notebook.	Generate new data	Digital	Experimental	.pdf	<1GB	
MS & CHROMATO-GRAPHY DATA	Data generated during LCMS, GCMS or other mass spectrometry techniques, or during analytical or separative HPLC runs. Data folders are 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.	Generate new data	Digital	Experimental	.raw, .pdf, .jpg	<1GB	
X-RAY DATA	Data generated during single crystal X-ray spectroscopy. The raw datasets are recorded by the responsible person on the X ray diffractometer. After processing, the cif files are provided to the researchers and uploaded to https://www.ccdc.cam.ac.uk in the framework of publications	Generate new data	Digital	Experimental	.cif	<1GB	

ELEMENTAL ANALYSIS	Elemental analysis data is generated by a CHNX combustion analyser. Data folders are generated by the combustion analyzer containing raw data and also all relevant acquisition parameters, stored as .dat files. Processed data is returned to the researcher on paper, and scanned to pdf and attached to the electronic lab notebook.	Generate new data	Digital	Experimental	.pdf	<1GB	
CHEMICAL COMPOUNDS	Products of chemical synthesis in powder or oil form, stored in vials. Vials contain mg to g quantities of purified product. Not all compounds are stored, mainly important intermediates and final products. Vials are stored in cryoboxes in Chem&Tech stock room 01.186, and metadata for each sample is stored in cheminventory cloud software.	Generate new data	Digital and Physical	Experimental	.xlsx	<1GB	up to four cryoboxes of 130x130x5cm.
SDS-PAGE GEL ANALYSES	Data generated by fluorescence scanning of gel-based experiments. Data	Generate new data	Digital	Experimental	.gel, .jpg	<1GB	
PROTEOMICS DATA	Data generated during LC-MS/MS analysis of samples at the SyBioMa facilities.	Generate new data	Digital	Experimental	.xlsx	<1GB	

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:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

The chemical biology experiments will make use of commercial human cell lines. Ethical approval will be sought for these tests but is still underway (CMT form E-2023-4162).

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- 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.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/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

2. 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 experimental data related to the project will be stored in a standardized DMP folder structure in OneDrive for Business. All experiments will be discussed in the Electronic Lab Notebook and will 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.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) 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.

3. Data storage & back-up during the research project

Where will the data be stored?

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.

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS.

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 estimated upper end of data volume in section 1 is <108GB. KU Leuven provides OneDrive storage of 2 TB for coworkers. 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?

Archival storage at KU Leuven is not publicly accessible and only people with permission (PI and/or his delegate responsible for data storage).

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

Archival data storage is centrally offered via KU Leuven at 270 Euro/TB/Year. This cost is covered by the research group.

4. Data preservation after the end of the research project

Which data will be retained for at least five 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 ELN data will be retained. All data collected in OneDrive during the course of the project will be retained in the archive storage.

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

The lab books will be stored in a dedicated cloud based ELN. Data will be exported to XML or PDF files at the end of the project for backup purposes and will also be stored on the on the university's central servers (with automatic back-up procedures). The compound characterisation data and reports will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

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

Archival data storage is centrally offered via KU Leuven at 270 Euro/TB/Year. This cost is covered by the research group.

5. Data sharing and reuse

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

- Yes, in an Open Access repository

All publications will be deposited in the LIRIAS repository for permanent storage already during the project. Also the data that are included in the supporting information of publications will be stored in the archive.

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

NA

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 in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

All data necessary to support the findings of research results will be published in supplementary information of the publications. No use will be made of data repositories. The data will be accessible for internal use in the archival storage managed by the project PI.

When will the data be made available?

Upon publication of the research results. The data relevant to the papers published will be disclosed via the supporting information section.

Which data usage licenses are you going to provide? If none, please explain why.

Publication data will be shared under license CC-BY-NC-ND-4.0.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

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

No additional costs are foreseen other than those for archival.

6. Responsibilities

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

The postdoctoral researcher is responsible for collecting all relevant data files and for entering the observational data in the ELN.

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

The postdoctoral researcher is responsible for storing all relevant data in OneDrive and Mbook.

Who will manage data preservation and sharing?

The PI (and/or a local data manager) is responsible for ensuring data preservation and reuse

Who will update and implement this DMP?

The postdoctoral researcher