Molecular mechanisms underlying peroxisome-mediated changes in the cellular hydrogen peroxide signaling network

A Data Management Plan created using DMPonline.be

Creators: Celien Lismont https://orcid.org/0000-0002-7388-8878, Marc Fransen https://orcid.org/0000-0001-9284-1197

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 1213623N

ID: 193351

Start date: 01-10-2022

End date: 30-09-2025

Project abstract:

For a long time, hydrogen peroxide (H2O2) was considered as a detrimental byproduct of oxidative metabolism. However, during the last decades, this molecule has moved into the forefront as a central second messenger in many biological processes. An important site of cellular H2O2 metabolism is the peroxisome. This fundamental research project is intended to improve our understanding of the molecular mechanisms underlying peroxisome-mediated changes in the cellular H2O2 signaling network. Specifically, I will (i) search for targets of peroxisome-derived H2O2 that are oxidized through a PRDX1-based redox relay, (ii) investigate whether peroxisomal matrix proteins are protected against H2O2-mediated oxidation by glutathionylation or CoAlation, (iii) determine the impact of disease-related disturbances in peroxisome function on the cellular sulfenome, and (iv) investigate how peroxisome-derived H2O2 modulates protein interactions within triorganellar membrane contact sites between peroxisomes, mitochondria, and the endoplasmic reticulum. Gaining more insight into these phenomena will advance our understanding of how peroxisomes serve as key regulators of H2O2-driven cellular processes.

Last modified: 20-01-2023

Molecular mechanisms underlying peroxisome-mediated changes in the cellular hydrogen peroxide signaling network
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
1	(Immuno)blots	Generate new data	Physical				350-500 blots
2	Scans of (immuno)blots	Generate new data	Digital	Experimental	.tif	<100GB	
3	(Immuno)fluorescence microscopy images	Generate new data	Digital	Experimental	.vsi, .tif, .jpeg	<100GB	
4	Mass spectrometry data	Generate new data	Digital	Experimental	.raw, .xls	<1TB	
5	Molecular biology and biochemistry experimental data	Generate new data	Digital	Experimental	.xls	<1GB	
6	Cell lines	Generate new data	Physical				15-20 cell lines; 3-10 tubes/cell line
7	Protein samples	Generate new data	Physical				2000-5000 samples; 1 tube/sample
8	Plasmid constructs	Generate new data	Physical				10-30 constructs; 3 tubes/construct

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 will not reuse existing data.

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.

- Yes, human subject data

This study will employ cell lines derived from commercially available human cell lines (Flp-In™ T-REx™ 293, Thermo Fisher Scientific, R78007; and HeLa, ATCC, CCL-2™). As such, questions regarding the collection, processing, disclosure of personal data, and the need for informed consent are not applicable. The Ethical Committee Research UZ/KU Leuven has approved an umbrella study protocol (reference: S63808) of the Department of Cellular and Molecular Medicine that (among others) contains all cell lines and procedures necessary to complete this project.

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).

1. All experiments performed in the wet lab, will be documented in laboratory notebooks. In addition, electronic files generated during the experiment (e.g., .xls, .tiff, .jpeg), figures, and overviews will be archived in a logical folder structure. The folder names will be meaningful, consistent, and allow co-workers to find relevant files easily. Files will be named according to the convention “short thematic description of the experiment_initials of the researcher_date (DDMMYYYY)_version number” (e.g., roGFP measurement_CL_18052023_v1). For every type of experiment, there will be a document containing all information (e.g., study design, experimental protocol, deviations from the protocol, other useful contextual information such as definitions of variables and units of measurement) necessary for a secondary analyst to use the data accurately and effectively.
2. Optimized protocols and manuscripts will be archived in separate folders following a logical folder structure. Note that, in addition to a detailed description of the experimental procedures in the manuscript, a copy of all primary data discussed in the manuscript will be saved in a subfolder for maximum traceability.
3. The generation and storage of cell lines and plasmid constructs are documented in the lab-specific database ConstructExpress which is accessible to all lab members.

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.

- Yes

1. **Microscopy** images will follow the *OME-XML* (Open Microscopy Environment eXtensible Markup Language) metadata standard. Of note, when exporting images as .tiff files, the microscopy software saves them in OME-TIFF format.
2. **Mass spectrometry** data and their accompanying metadata will be made publicly available in the PRoteomics IDEntification (PRIDE) database. This database is compliant with the *MIBBI* (Minimum Information for Biological and Biomedical Investigations) metadata standard.
3. To provide structure, context, and meaning of the **other primary data** that will be collected, we will use (1) *descriptive metadata* relevant for the findability and interpretation of data (e.g., title, abstract, author, keywords, cell type, dose, exposure, etc.), (2) *structural metadata* facilitating the navigation of the electronic data (e.g., structuring tags such as page numbers, title, table of contents, sub-object relationships, etc.), and (3) *administrative metadata* that indicate by who, when, and why the data were created. The latter data will also contain technical (e.g., filetype, date, size, etc.) and purely administrative (e.g., collaborators, etc.) properties.

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

Where will the data be stored?

1. Original hard-copy data (e.g., immunoblots) will be stored in laboratory notebooks.
2. Primary mass spectrometry data (.raw files) are initially stored on the computer attached to the LC-MS/MS device and backed up on a separate computer for data analysis at the proteomics core facility. After publication, all data related to the published project will be transferred to 2 external hard drives that will be stored in 2 separate locations (Marc Fransen's office and the microscopy room). In addition, published data will be made available in the PRIDE database.
3. Other electronic data files (e.g., scans of immunoblots, fluorescence microscopy images, spreadsheets, etc.) will be kept in our research unit folder (J:\GBW-0071_LIPIT) on the university's ICTS server. Copies can be made and kept on personal devices.
4. Physical samples (cell lines, plasmid, and protein samples) will be stored in liquid nitrogen, at -80°C, or -20°C, depending on the nature of the sample.

How will the data be backed up?

1. Most data will be stored on the university's ICTS server (J:\GBW-0071_LIPIT) with automatic daily back-up procedures.
2. The two external hard drives that contain the raw mass spectrometry data serve as copies of each other. Each time data are added to one drive, they will immediately be backed up in the other drive.
3. Of each cell line, at least 3 aliquots will be frozen in liquid nitrogen at any given time. During the course of the project, these aliquots will be moved to the university's cryofacility where temperature is constantly monitored.
4. Plasmids are backed up as glycerol stocks of transformed bacteria and stored at -80°C. This freezer is equipped with a temperature monitor. When the temperature rises above -65°C, the university's central dispatch contacts Marc Fransen and/or Celien Lismont.
5. Other samples (e.g., protein samples) are meant to be consumed rather than backed up.

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

1. Currently, there is still ±400 GB storage capacity available in our ICTS server folder (J:\GBW-0071_LIPIT). We anticipate to add 40-100 GB onto this server in the course of this project. Nevertheless, the capacity might not suffice for all ongoing projects in the laboratory. When needed, we will contact the ICTS service desk to increase our research unit folder storage capacity.
2. The external hard drives each have more than 1.5 TB of unused capacity. This will more than suffice for all ongoing projects in the lab.
3. Storage space in the cryofacility and the laboratory freezers suffices for this project.

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

1. We will take care that only authorized users (e.g., team members involved in the project) can access the data through specific user accounts. Importantly, these accounts will be disabled from the moment these team members are no longer involved in the project.
2. The external hard drives are only accessible by team members and other persons that have access to the office of Marc Fransen and/or the microscopy room. As these external drives do not contain any personal data, it concerns projects that have already been published, and the interpretation of the data is (near to) impossible without additional information on the experimental conditions, we are convinced that this course of action offers sufficient protection in case of theft. In addition, both locations are separated by a fire door, providing adequate protection against loss of data by fire.
3. Both the research building and the cryofacility are access-controlled with a badge system.

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

1. In view of the expected size (13-33 GB per year) of the data that will be stored and the estimated costs (1 Euro per GB and per

year), the maximum cumulative costs of data storage and back-up during the course of the project is estimated to be 200 Euro. The host laboratory has the financial means to pay for the data storage.
2. As the cryofacility is a new service and we are currently in a transition fase, prices have not yet been determined. As soon as possible, this DMP will be updated with this information.

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...).

1. All hard copies and electronic data will be retained for a period of at least 10 years after the end of the project, conform the KU Leuven RDM policy.
2. Likewise, plasmids and cell lines will be stored for at least 10 years. Note that these resources are likely to be reused within the research laboratory or by others (e.g., upon request after publication).
3. Other sample types (e.g., protein samples) are ordinarily consumed during the experiments and will therefore not be available for long-term storage.

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

1. All hard copy data will be stored in the office of Marc Fransen.
2. The external hard drives containing all primary mass spectrometry data will be stored in the office of Marc Fransen and in the microscopy room.
3. Other electronic data files (e.g., scans of immunoblots, fluorescence microscopy images, spreadsheets, etc.) will be kept in our research unit folder (J:\GBW-0071_LIPIT) on the university's ICTS server.
4. Publications and published mass spectrometry datasets will be available through Lirias and the PRIDE database, respectively.
5. Plasmids (bacterial glycerol stocks) and cell lines will be stored at -80°C in the laboratory and in liquid nitrogen at the university's cryofacility, respectively.

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

1. In view of the expected size (maximum 100 GB) of the data that will be stored, the estimated cost (1 Euro per GB and per year) will be at maximum 100 Euro per retention year.
2. The costs related to the use of the cryofacility will be calculated once the prices are announced.

We trust that the insights gained from this versatile and innovative project lead to successful funding applications that will enable us to cover the data storage costs.

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
- Other, please specify:

1. Datasets that underpin publications will be made freely available in Open Access Repositories.
2. Plasmids and cell lines will be made available upon request.
3. Unpublished data may be shared.

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

1. Given that the data will be deposited in Lirias (a KU Leuven Association-based open access repository and information archiving system) and the PRIDE database, they will be freely accessible to all.
2. Plasmids and cell lines will be shared after publication with all academic researchers that request access.
3. Relevant unpublished data may be shared in the context of a collaboration.

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.

1. Datasets that underpin publications will be made freely available as (i) supplementary information within the corresponding Lirias-archived publications or (ii) in the PRIDE database.
2. Electronic data shared in the context of a collaboration can be distributed by email or a cloud service such as Dropbox, depending on the size of the data.
3. Plasmids and cell lines will be shipped or can be picked up upon request.

When will the data be made available?

1. Publications (including supplementary data) will be uploaded in Lirias 2.0 by latest at the date of publication. However, depending on the publisher, there may be an embargo period of maximally 6 months (see Federal Bill (Article XI. 196 $2/1) that is in force since 18 September 2018).
2. Proteomics datasets will be uploaded into the PRIDE database before the review process with restricted access for the reviewers only and made freely accessible upon acceptance for publication.
3. Plasmids and cell lines can be requested upon publication of the manuscript.
4. Within the context of a collaboration, data can be shared at any favorable time point.

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

1. Publications in Lirias 2.0 will have a 'public access (as soon as legally possible, verified by open access Support Desk)' license. This means that the full text will be openly accessible as soon as the publisher's open access policy and the Belgian open access legislation allow.
2. Proteomics datasets submitted to the PRIDE database fall under PRIDE's license agreement (https://www.ebi.ac.uk/pride/markdownpage/license).
3. When sharing other data, we will count on the legal services of KU Leuven Research & Development to draft suitable material transfer agreements.

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.

- Yes

Proteomics datasets in PRIDE receive a PX identifier PXD030782 which will be included in the relevant publication.

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

1. The deposition of pre-print versions of publications in Lirias and proteomics datasets in PRIDE is free of charge. The costs related to data that still need to be published in year 4 (= within one year of the end of the project), are foreseen in the budget.
2. Shipping of physical materials (e.g., cell lines) is financed by the receiving researcher.

6. Responsibilities

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

All people involved in the project (Marc Fransen, Celien Lismont, bachelor/master/PhD students, technicians) will be responsible for the documentation and verification (e.g., accurateness) of primary source data and metadata.

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

Celien Lismont will be responsible for data storage, and staff from the centralized ICT service of KUL will be responsible for backing up the data on a daily basis.

Who will manage data preservation and sharing?

All team members will be involved in the data curation activities, which will enable data discovery and retrieval, maintain quality, add value, and provide for data re-use over time. Marc Fransen will be responsible for ensuring data preservation and reuse.

Who will update and implement this DMP?

Celien Lismont and Marc Fransen share responsibility to update and implement this DMP.