# Structural dynamics of secretory protein sorting, targeting and translocation Acronym: DynamicSTarT

*A Data Management Plan created using DMPonline.be*

**Creators:** Lily Karamanou, n.n. n.n.

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** n.n. n.n.

**Data Manager:** Lily Karamanou

**Grant number / URL:** G009423N

**ID:** 200721

**Start date:** 01-01-2023

**End date:** 31-12-2028

**Project abstract:**
So far, our understanding of protein folding has been fragmented: folding events upon ribosomal exit, the structural basis of chaperone/preprotein interaction, folding regulation under stress, the role of periplasmic chaperones, folding of porins coupled to outer membrane integration. However, all proteins possess intrinsic dynamics properties that affect their folding and their intrinsic motions once folded. The intrinsic dynamics of chaperones have been correlated with three distinct functions: unfold – hold – refold. Secretory clients have evolved to take advantage of these properties to different extents depending on their own internal structural features. By using exclusively secreted proteins and chaperone examples, DynamicSTarT will probe a fundamental question: how do the dynamics of preprotein clients and chaperones underlie the complete ribosome-cytoplasm-translocase route.

**Last modified:** 28-06-2023

# Structural dynamics of secretory protein sorting, targeting and translocation
## Acronym: DynamicSTarT
## Application DMP

---

## Questionnaire

**Describe the datatypes (surveys, sequences, manuscripts, objects … ) the research will collect and/or generate and /or (re)use. (use up to 700 characters)**

New data will be collected as well as own, existing data will be used.
*Genetic constructs and bacterial strains:* Wild type or mutant genes that will be generated for this study (by subcloning or chemical synthesis) will be preserved as plasmid preps at -20°C, or after transformation in DH5alpha cells, as frozen bacterial glycerol stocks at -80°C. Genetic constructs will be sequenced (.txt; ~400 KB total) and described using Vector NTI (.gb; ~10 MB total). Bacterial strains, preserved as bacterial glycerol stocks at -80°C will be described in FileMaker pro (.fmp; ~2.5GB total).
*Global and local HDX-MS data:* m/z digital spectra of protein samples (.DAT; ~1-1.5 TB total).
*MD simulation data:*(.pdb, .xtc; ~2-5TB)
*Single molecule FRET data*:(.ptu, .csv; ~0.5TB)

**Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)**

1. Designation of responsible person (If already designated, please fill in his/her name. Dr L Karamanou (Research Manager)
2. Storage capacity/repository
   - during the research: All data under analysis or unpublished are stored on 3 local servers that are password protected, not connected to the internet and on KU Leuven/OneDrive servers. Servers are backed-up once per month in 2 copies, one on an external hard disk locked on the PI's office, the other one on the Lab DropBox account (password protected).
   - after the research: Data will be retained for at least 5 years on KUL server space (archive). Two additional copies on external hard disks will be made (one remains with the PI, the second one with the researcher). Descriptions are filed in our lab data bases (FileMaker Pro Advanced) and stored (these files are updated and backed-up once per month).

**What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)**

Not applicable

**Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)**

Not applicable

**Which other issues related to the data management are relevant to mention? (use up to 700 characters)**

For security reasons (but not only; see below), many lab computers are not connected to the internet therefore data are transferred from/to them manually.
Some of these computers run specialized software from vendors that are very slow in updating and oftentimes they stay behind the current version of windows. Updating them systematically cripples the specialized software so we can only update them when instructed by the vendor.

# Structural dynamics of secretory protein sorting, targeting and translocation
## Acronym: DynamicSTarT
## 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.**

| Dataset Name | Description | New or reused | Digital or Physical | Only for digital data<br>Digital Data Type | Only for digital data<br>Digital Data format | Only for digital data<br>Digital data volume (MB/GB/TB) | Only for physical data<br>Physical volume |
|---|---|---|---|---|---|---|---|
| S1 | Genetic constructs | Generate new data | Physical | NA | *NA* | *NA* | Glycerol stocks of bacteria stored at -80C<br><100 |
| S2 | HDX-MS data | Generate new data (plus already accumulated preliminary unpublished data) | Digital | Experimental | .csv, .xml | <5TB | NA |
| S3 | smFRET data | Generate new data (plus already accumulated preliminary unpublished data) | Digital | Experimental | .csv .ptu | <0.5TB | NA |
| S4 | MD simulation data | Generate new data | Digital | Simulation | .pdb .xtc | <5TB | NA |
| | | | | | | | |

**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 plan to use preliminary own data (unpublished) eg; genetic constructs/strains; HDX-MS and smFRET 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.**

- No

**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.**

- Yes

Biomedical/biotechnological applications are anticipated and will be explored as appropriate: a. novel antibiotics that block secretion (ongoing collaboration with the medicinal chemistry company CD3 of KUL), b. optimized production of secreted recombinant proteins, c. improved understanding of protein aggregation/chaperone interactions.

**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).**

Protocols are saved for each WP, as word files. HDX-MS and smFRET data collections are summarized by README.txt files that describe the procedures used during data collection and summarize the experimental conditions. Genetic constructs / strains are filed with their descriptions using Filemaker Pro databases and word files are used to summarize them.

**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

If needed the RDR metadata format will be used

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

**Where will the data be stored?**

Data under analysis or unpublished are stored on 3 local servers that are  password-protected, not connected to the internet, and on paid storage (KU Leuven/OneDrive) servers.

**How will the data be backed up?**

Servers are backed-up once per month in 2 copies, one on an external hard disk locked on the PI's office, the other one on the Lab DropBox account (password protected).

**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

-Every KU Leuven employee and student has their own OneDrive with 2 TB (or 2000 GB) of storage space.
-Our local servers have 5-10TB storage space.
-The lab provides each lab member with an external disk of 5-10TB capacity to back up all of the data the user produces.

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

Once per month all servers are backed-up in 2 copies, one on an external hard disk locked on the PI's office, the other one on the Lab DropBox account (password protected).

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

-Onedrive is free for KUL members.
-The lab provides for the servers (plus maintenance) and the external hard disks from lab funds (other that the current FWO project).

## 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 of the data will be preserved for a minimum of 5 years.

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

-In external hard disks locked in the PI's office.
-In the lab Dropbox account.
-In paid KUL storage space (Archive).

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

~50-200 euros/year. For archiving data with the K-Drive 50% of the cost is covered by the group of Biomedical sciences. The rest of the cost will be covered by lab maintenance funds.

## 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.**

- Other, please specify:

Published data will be available to users upon request.

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

The users will have to declare the project/reason/purpose that the data will be used to/for (to secure no conflict with the lab interests) and they will have to credit/acknowledge the lab for using previously accumulated data.

**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.**

- Yes, Other

There is no problem in sharing the data as long as there is no conflict of interest with an ongoing lab project and as long as the lab is credited for the data collection.

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

Via the KU Leuven repository (RDR)

**When will the data be made available?**

After publication of our study.

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

A creative commons attribution-nonCommercial-ShareAlike 4.0 International license (CC-BY-NC-SA 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.**

- Yes

A DOI will be provided in the future.

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

RDR use is free for KUL labs, therefore no cost is expected

## 6. Responsibilities

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

Lily Karamanou (the lab's research manager) will be responsible, with help from the research fellows involved in the project and following consultation with the PI (Prof. T. Economou)

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

Lily Karamanou (the lab's research manager) will be responsible, with help from the research fellows involved in the project and following consultation with the PI (Prof. T. Economou)

**Who will manage data preservation and sharing?**

Lily Karamanou (the lab's research manager) will be responsible, in consultation with the PI (Prof. T. Economou)

**Who will update and implement this DMP?**

Lily Karamanou (the lab's research manager) will be responsible, in consultation with the PI (Prof. T. Economou)

# Structural dynamics of secretory protein sorting, targeting and translocation Acronym: DynamicSTarT
# GDPR

## GDPR

**Have you registered personal data processing activities for this project?**

- Not applicable

Created using DMPonline.be. Last modified 28 June 2023

7 of 8

## DPIA

**Have you performed a DPIA for the personal data processing activities for this project?**

- Not applicable