Plan Overview

A Data Management Plan created using DMPonline.be

Title: Developing XNA aptamer selection tools for the advancement of aptamer diagnostic and therapeutic applications.

Creator: Saphyra Thonon

Principal Investigator: 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.

Project abstract:

Aptamers are short, single-stranded oligonucleotide sequences that bind with high affinity and specificity to a target; e.g. a small molecule, a biomarker or a protein. Together with numerous beneficial characteristics, aptamers have held great promise as therapeutics, (point-of-care) diagnostics, selective drug delivery systems and biomarker discovery platforms in a wide range of pathologies.

However, thus far aptamer selection has been time-consuming, laborious and unpredictable. Additionally, the aptamer definition has been vague, leading to publications of subpar aptamers. Consequently, there is a big gap between fundamental aptamer research and clinical application. To bridge this gap, perspectives in the field need to change, and more efficient selection tools must be developed.

With this project, I aim to improve the XNA aptamer selection process for clinical application by making aptamer selections more robust, more predictable and faster. I will achieve this by adapting previously established tools, and by engineering a novel DNA-dependent XNA transcriptase; using the selection of an XNA aptamer against human Interleukin 2 as a proof-of-principle.

The knowledge gained by this project will enhance future XNA aptamer research avenues through faster and more efficient selections, exploring the potential of different XNA chemistries and further narrowing the gap between research and clinical applications with numerous valorisation opportunities.

ID: 211140

Start date: 01-11-2024

End date: 31-10-2028

Last modified: 17-04-2025

Developing XNA aptamer selection tools for the advancement of aptamer diagnostic and therapeutic applications.

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	
Protocols & SOPs		Generate new data & reuse existing data	Physical & digital	Other	.pdf .docx	<100MB	<50 pages
DNA, RNA and XNA	Purchased oligos, nucleic acid libraries and purified plasmids	Generate new data & reuse existing data	Physical				<1000 1.5 mL tubes
Purified proteins	Purified T7 variants, XNA polymerases and aptamer targets	Generate new data & reuse existing data	Physical				<500 1.5 mL tubes
Glycerol stocks	Libraries and protein constructs	Generate new data & Reuse existing data	Physical				<500 1.5 mL tubes
DNA, RNA, XNA and protein amino acid sequences	Oligos, libraries, plasmid maps, protein nucleic acid sequences, published protein variants	Generate new data & reuse existing data	Digital	Experimental Other	.dna .fa .pdf .jvp .txt .xls	<100GB	
Alignment results	T7 Multiple Sequence Alignment	Generate new data	Digital	Experimental	.fa .fasta .aln	<1GB	

Protein & nucleic acid structures	T7 polymerase structures and nucleic acid structure predictions	Reuse existing data	Digital	Experimental Simulation data	.pdb	<1GB	
Electrophoretic gel scans	Agarose, SDS-PAGE, UREA-PAGE	Generate new data	Digital	Experimental	.gel .png .jpg	<1GB	
Fluorescence read-outs	Activity assays, dPCR, concentration measurements	Generate new data	Digital	Experimental	.xls	<1GB	
Affinity measurements	Fluorescence polarisation, SPR	Generate new data	Digital	Experimental	.xls	<100GB	
Sangar DNA sequencing results		Generate new data	Digital	Experimental	.ab1 .pdf .txt	<100GB	
NGS analyses		Generate new data & reuse existing data	Digital	Experimental	.py .fasta .fastq	<100GB	
NMR spectra	Aptamer-protein structures	Generate new data	Digital	Experimental	ser, acqu, acqus, proc, procs, pdata folder, procpar, .jdx .txt . csv .xml .cara	<100 GB	

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:

Protocols & SOPs: protocols delivered in frequently used kits or existing standard lab protocols.
DNA: plasmids containing the T7 RNA polymerase previously cloned in the lab and oligos previously used in the lab for aptamer selection.
Glycerol stocks and purified proteins: T7 RNA polymerase variants previously cloned by Prof. Pinheiro's lab as well as aptamer selection targets.
DNA and protein amino acid sequences: existing sequences from GenBank, Addgene, Uniprot, published literature as well as existing nucleic acid libraries previously used by Prof. Pinheiro's lab.
Protein & nucleic acid structures: published predicted and determined 3D structures of proteins (e.g. T7 RNA polymerases) and nucleic acids (aptamers).
NGS analyses: NGS results will be analyzed using the Python scripts established by Prof. Pinheiro's lab.

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

The T7 XNA Transcriptase has the potential for commercial valorization. The KU Leuven Research & Development (LRD) will be contacted to discuss patenting and licencing opportunities.

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 experiments are documented in an electronic lab notebook using Benchling by the performing scientist following standard operating procedures (SOPs), which have been or will be written down. Each notebook contains the background and rational with the objective, protocols, materials and samples used, results and conclusions, linking to other notebooks as needed.
Typically, software used to generate experimentel data (e.g. fluorescence read-outs, NMR data) automatically generate files containing the raw data and accompanying experimental details. The raw data of each experiment will be sorted per experiment type and stored in separate folders.
Processed data will be stored in separate folders with the same name containing links to their respective raw data files. Separate documents of non-experimental nature will be sorted and stored in a documents folder. Data will be accompanied by README.txt files explaining the design/protocol, analysis methods, results, labels used and references to the electronic lab notebook.
Metadata will link the data files, lab samples, and experimental notes (including descriptions of equipment, setting, and used experimental settings).
New folders will be created for processed data with links to the raw data included when needed for specific publications.

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

When depositing data in a local or public repository, the final dataset will be accompanied with a README.txt file containing all the relevant information, following the Dublin Core Metadata standard. 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.

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

Where will the data be stored?

Experimental raw data (e.g. fluorescence read-outs, NMR data) are typically automatically saved to the local network by the software. All data will be stored on OneDrive for business (KU Leuven) and Syncthing will be used for additional back-ups and long-term storage.

How will the data be backed up?

Through OneDrive-client, automatic back-ups are made of local files to OneDrive.
Syncthing is being used to create an automated backup service of local files (linked to OneDrive or additional local files when necessary) both onsite and offsite, relying on personal machines from Prof. Pinheiro. This creates redundancy with data stored in multiple secure locations.

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 expected data volume falls within the free data storage allocation.

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

OneDrive for business is secured using the KU Leuven data security systems and authentication.
Syncthing creates a peer-to-peer encrypted network that must be authorized by both machines via a randomly-generated 128-bit key. Both ends must authorize the sharing of a folder before syncthing will allow data transfer. The network itself was set up in the Pinheiro lab as a spoke and hub model, with the spoke (individual researchers) only able to deposit data into the hub (onsite). The hub is backed up to an offsite secure server.

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

Both OneDrive for business (for KU Leuven based researchers) and Synthing are free to use.

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 data will be preserved for at least 10 years according to KU Leuven RDM policy.

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

KU Leuven RDR and the Open Science Foundation

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

The expected data volume falls within the free data storage allocation.

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

All data will be kept open as much as possible, apart from datasets (NGS, DNA sequences, etc.) that contain data with potential commercial value.

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

Restricted data - those with commercial value - may be shared with suitable MTAs.

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.

KU Leuven RDR and the Open Science Foundation.

When will the data be made available?

Data will be made available upon publication of research results.

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

- CC-BY 4.0 (data)
- Data Transfer Agreement (restricted data)
- MIT licence (code)

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 PID will be added upon deposit in a data repository.

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

The expected costs are negligable.

6. Responsibilities

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

The researcher will manage data documentation and metadata during the project.

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

Both the researcher (primary) and supervisor (secondary) will manage data storage and backup during the project.

Who will manage data preservation and sharing?

The researcher will manage data preservation and sharing during the project. After the completion of the project, the supervisor will manage the long term data preservation and sharing.

Who will update and implement this DMP?

During the project, the researcher will update and implement this DMP, whilst the supervisor will carry the end responsibility for this DMP.