

---

# Computer-aided design, selection and optimisation of multi-enzyme biocatalytic processes for predictive industrial biomanufacturing

*A Data Management Plan created using DMPonline*

**Creator:** Will Finnigan

**Affiliation:** University of Manchester

**Funder:** Biotechnology and Biological Sciences Research Council (BBSRC)

**Template:** BBSRC Template Customised By: University of Manchester

## **Project abstract:**

In order to tackle the challenges of engineering biological systems for future industrial biomanufacturing, new computational tools will be required which aid in the design of multi-enzyme cascade processes. For example, (A) computer-aided pathway design tools encompassing both biocatalytic and chemocatalytic transformations are essential for the design of industrially relevant routes, yet none currently exists. In addition, (B) algorithms for mapping protein sequence to function to allow us to intelligently select the best biocatalysts during route planning will be essential for success. Finally, (C) the use of mathematical modelling will need to be integrated into the design process to help us rapidly optimise the increasingly complex multi-enzyme systems being developed. This proposal addresses all these aspects with the overall aim of achieving predictable in-silico design of biocatalytic and industrial biomanufacturing processes.

**ID:** 76733

**Start date:** 01-05-2022

**End date:** 30-04-2025

**Last modified:** 29-04-2021

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# Computer-aided design, selection and optimisation of multi-enzyme biocatalytic processes for predictive industrial biomanufacturing

---

## Manchester Data Management Outline

**1. Will this project be reviewed by any of the following bodies (please select all that apply)?**

- Funder

**2. Is The University of Manchester collaborating with other institutions on this project?**

- Yes - Part of a collaboration and not handling data

Working with AstraZeneca, Novartis and EnginZyme

**3. What data will you use in this project (please select all that apply)?**

- Re-use existing data (please list below)
- Acquire new data

**4. Where will the data be stored and backed-up during the project lifetime?**

- University of Manchester Research Data Storage Service (Isilon)

**5. If you will be using Research Data Storage, how much storage will you require?**

- 1 - 8 TB

**6. Are you going to be receiving data from, or sharing data with an external third party?**

- No

**7. How long do you intend to keep your data for after the end of your project (in years)?**

- 5 - 10 years

***Guidance for questions 8 to 13***

Highly restricted information defined in the [Information security classification, ownership and secure information handling SOP](#) is information that requires enhanced security as unauthorised disclosure could cause significant harm to individuals or to the University and its ambitions in respect of its purpose, vision and values. This could be: information that is subject to export controls; valuable intellectual property; security sensitive material or research in key industrial fields at particular risk of being targeted by foreign states. See more [examples of highly restricted information](#).

Personal information, also known as personal data, relates to identifiable living individuals. Personal data is classed as special category personal data if it includes any of the following types of information about an identifiable living individual: racial or ethnic origin; political opinions; religious or similar philosophical beliefs; trade union membership; genetic data; biometric data; health data; sexual life; sexual orientation.

Please note that in line with [data protection law](#) (the UK General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de-identified) as soon as practically possible. You must obtain the appropriate [ethical approval](#) in order to use identifiable personal data.

**8. What type of information will you be processing (please select all that apply)?**

- No confidential or personal data

**9. How do you plan to store, protect and ensure confidentiality of the participants' information (please select all that apply)?**

- Not applicable

**10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?**

- Not applicable

**11. Will the participants' information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?**

- Not applicable

**12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?**

- Not applicable

**13. Are you planning to use the personal information for future purposes such as research?**

- No

**14. Who will act as the data custodian for this study, and so be responsible for the information involved?**

William Finnigan

**15. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).**

2021-04-29

## **Data areas and data types**

## **Outline the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images**

Computer code will be generated during the project, primarily in the form of python, javascript, html or css files. Computer code will be made available as open-source software where possible. Further data will be generated in the form of nucleotide and protein sequences, plasmid maps, culture and enzyme preparation protocols, assay and screening protocols, and biotransformation protocols, chromatographic and spectroscopic data. In order to permit long-term accessibility, data will be stored in open, non-proprietary formats commonly used within the scientific community.

## **Standards and metadata**

### **Outline the standards and methodologies that will be adopted for data collection and management, and why these have been selected**

I will be responsible for data collection, as well as data management and ensuring the quality of data collection methods used through the research programme. Standardized protocols will be established for data collection procedures to ensure the consistency and reliability of the data. Experimental methodology and data collection methods will be reviewed by experienced colleagues to guarantee the highest standards are met. Detailed descriptions of experimental work will be safely stored and disseminated through publications and conference presentations.

## **Relationship to other data**

### **State the relationship to other data available in public repositories**

Where appropriate, data such as enzyme protein sequences will be linked to the UniProt or NCBI databases.

## **Secondary Use**

### **Outline the further intended and/or foreseeable research uses for the completed dataset(s)**

Data will be placed in the 'RetroBioCat' database.

## Methods for data sharing

**Outline the planned mechanisms for making these data available, e.g. through deposition in existing public databases or on request, including access mechanisms where appropriate**

Computer code will be stored in a code repository such as GitHub and made publicly available where appropriate. Any and all novel sequence or structural data collected as part of this project will be deposited to the appropriate public databases, (e.g. GenBank, EMBL, Protein Data Bank, etc.). All plasmids and strains developed under this Fellowship will be archived and made available to share by request after completing an appropriate Materials Transfer Agreement. Experimental methods and technical data will be presented in publications as part of either the main text or the supporting information, as well as at conferences. Articles published with a separate document for supporting information will contain instructions on how to access the supporting file. A statement regarding our data sharing policy will also be included in all publications that result from this project. Any requests for additional data can be made to the Fellow, who will arrange the transfer. The Fellow will also pursue options for making articles freely available on the publisher's website through Open Access, according to institutional guidelines.

## Proprietary data

**Outline any restrictions on data sharing due to the need to protect proprietary or patentable data**

Patentable data, methods and/or technologies will be secured through the University's intellectual property commercialization department and technology transfer office (UMI3 and UMIP) as soon as possible to reduce any delay in the submission of publications or data release.

## Timeframes

**State the timescales for public release of data**

Data generated during the project will be released in line with established practices within the relevant research communities. Data to be made available through scientific publications will be compiled and submitted in a timely manner, as soon as possible after all necessary data has been collected and analysed. According to the University of

Manchester Research Data Management Policy, any relevant data that are likely to have long term value, such as data that substantiate research findings or represent records of the University, will be curated and maintained after the completion of the project for as long as they remain valuable.

## **Formats**

### **State the format of the final dataset**

Data will be stored in standard formats, depending on the type of data.