

DATA MANAGEMENT PLAN

Roles and Responsibilities.

Calvin Henard, Ph.D., Principal Investigator will be responsible for supervising all aspects of this proposal. He will supervise and train the research assistant and graduate student who will perform the strain development and analyses, and microbe cultivations. A postdoctoral fellow will be co-mentored by Drs. Henard and Alonso that will perform the ^{13}C tracing experiments and analyses. As the project leader, Henard will be responsible for recording data, writing reports for the agency where needed, and dissemination of data via presentations and publications. Henard will be the primary point of contact for Agile BioFoundry communication and interactions. Further, Henard will ensure that personnel are appropriately trained in data recording and management.

Henard has a diverse microbiological background working with an array of prokaryotic and eukaryotic microbes and has extensive experience in bacterial genetics and microbial metabolism. Henard has been studying methanotroph biology for several years and has become a recognized expert in the field via his contributions to the expansion of methanotroph genetic tools and methodology and metabolic engineering in these organisms.

Types of Data. All the data collected under this research project will be processed and stored in a systematic manner. Data collection title, location, and procedures are all recorded using the Benchling electronic laboratory notebook and archived within the PI's research group cloud. All digital data and documents are time stamped and traceable. Data is also stored and backed-up on a shared drive managed by the UNT College of Science as well as within the research lab. We will generate:

1. ^{13}C -labeling data of intermediary metabolites and biomass: mass isotopomer labeling information by both gas chromatography-mass spectrometry (GC-MS) and liquid chromatography tandem mass spectrometry (LC-MS/MS);
2. mathematical models;
3. metabolic maps of central metabolism: graphical representation of carbon flux maps using Omix Viewer;
4. several plasmid constructs;
5. raw sequencing data;
6. *E. coli* strains and genetically modified *M. capsulatus* strains/biocatalysts;
7. Genome-scale metabolic model

Data and Metadata Standards. The ^{13}C -labeling information will be stored as ".raw" files for mass spectrometry data because these formats are standard and readable by the majority of free and licensed software. Plasmid constructs are designed in silico and stored as genbank (.gb) files. Quantitative biochemical and metabolite data will be stored in ".csv" files, which are recognized by an array of statistical software. Any custom computer code generated during these investigations will be deposited in the Henard laboratory GitHub database or publicly available databases. All data generated from research under this award will be stored in a data repository on network hard drives. Data will be stored up to and after any publication on secure network drives and will be maintained for a minimum of five years after the award has closed. The Agile BioFoundry collaborators will follow DOE policy of data management and dissemination.

Policies for access and sharing and provisions for appropriate/privacy. UNT agrees to manage the intellectual property (IP) generated because of this project in a manner that promotes the dissemination of project results, and timely publication of research discovery and knowledge to the scientific communities. Discussions regarding IP will be initiated as the need arises during the project with the DOE ABF and UNT. Ownership of specific IP will be determined in accordance with United States of America law regarding invention and authorship and DOE policy. Any genetic tools applicable to the research community developed herein will be deposited at Addgene repository following publication. Further, *M. capsulatus*

genetically modified strains will be available upon request following publication. In accordance with the DOE Open Access Policy, all research data displayed in publications resulting from this proposal will be made available in open, machine-readable, digitally accessible format for free and open access to the public. Publicly available datasets will be identified with an OSTI-supplied digital object identified (DOI).

Dissemination of Data. Publications in high impact journals indexed through PubMed and Web of Science will continue to be a major aspect of our information sharing strategy. To facilitate the dissemination of the information, we have requested a small budget to partially cover journal fees for free online access to all publication related to the proposed research. Further, data and models developed through this project will also be deposited to existing databases including, but not limited to, the DOE Systems Biology Knowledgebase (KBase).