Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D17D0CBCE5

Title: Development of Cutibacterium-specific immunoassays to identify true Cutibacterium acnes infections.

Creator: William Mccoy iv - **ORCID:** 0000-0001-5115-3793

Affiliation: Washington University in St. Louis (wustl.edu)

Contributor: Seonyoung Kim

Funder: National Institute of Allergy and Infectious Diseases (niaid.nih.gov)

Funding opportunity number: PA-20-185

Template: NIH-Default DMSP

Project abstract:

More than one million infections occur yearly in the USA due to an indwelling medical device (IMD), like a catheter, pacemaker, mechanical heart valve, cerebrospinal fluid (CSF) shunt, or artificial joint. These infections incur significant medical costs, morbidity, and mortality (>25% for mechanical heart valves), and microbes that live on or in the human body are the most often cause of these infections. Infected devices frequently must be replaced while undergoing antibiotic therapy due to the presence of microbial biofilms that shelter microbes from antibiotics. Strategies to prevent these infections (standard surgical skin preparation with topical anti-microbial agents like ChloraPrep, DuraPrep, and povidone-iodine scrub; prophylactic antibiotics) do not effectively control Cutibacterium acnes, a common cause of IMD infections. C. acnes is a normal inhabitant of human skin, where it grows as a biofilm. Widespread antibiotic use for acne vulgaris has led to significant C. acnes antibiotic resistance. C. acnes is now the most common cause of shoulder periprosthetic joint infection (PJI) and causes a high percentage of sternotomy (34%) and CSF shunt infections (6%). The indolent nature of C. acnes infections (minimal/no erythema, drainage, fever; normal labs) frequently delays their workup. It then takes 1-2 weeks to culture C. acnes anaerobically. True C. acnes infections are also difficult to differentiate from environmental contaminants. An assay is needed to quickly identify C. acnes-specific growth in humans. Our lab has generated antibodies that recognize a potential C. acnes growth biomarker and used them to

develop assays. We propose to determine what (AIM 1) our antibodies bind on the biomarker, (AIM 2) type of C. acnes growth produces this biomarker, and (AIM 3) effect ligand-binding has on our antibodies' ability to detect this biomarker. Our study will help to develop the first diagnostics specific for C. acnes growth, and these assays will allow for rapid (<24 hours) identification of true C. acnes infections leading to improved patient outcomes (decreased morbidity/mortality) and lower healthcare costs.

Start date: 04-01-2025

End date: 03-31-2030

Last modified: 07-08-2024

Copyright information:

Development of Cutibacterium-specific immunoassays to identify true Cutibacterium acnes infections.

Data Type

Types and amount of scientific data expected to be generated in the project: *Summarize the types and estimated amount of scientific data expected to be generated in the project.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This study will generate microbial genomic data only to confirm the sequence of plasmid or reference microbe strain. Therefore, **this project is not within the National Institutes of Health (NIH) Genomic Data Sharing (GDS) policy's intended scope.**

No studies in this proposal will utilize specimens collected from patients. Therefore, **this study is not considered human subjects research.**

This project will produce the following data types and amounts.

Table 1: Data types and amount

Specific Aim	Subaim	Data Type	Data/File Format	File Size*	File Number*	Total Data
1	A-D	Protein Gel	.TIFF	<10MB	<500 files	<50GB
1	A-D	Western Blot (WB)	.TIFF	<10MB	<100 files	<1GB
1	B	X-ray Diffraction Data	.TIFF, .CIF, .mmCIF, .pdb	<10MB	<2000 files	<20GB
1	C	Binding Data	.csv	<10MB	<500 files	<5GB
1	A,B	Plasmid Sequencing**	.fasta	<10MB	<50 files	<1GB
1	A,B,D	ELISA	.csv	<10MB	<500 files	<5GB
1	B,E	Mass Spectrometry (MS)	.csv	<10MB	<100 files	<1GB
1	D,E	Chromatography	.csv	<10MB	<100 files	<1GB
1	E	Electron Microscopy (EM)	.TIFF	<2000MB	<9000 files	<18000 GB
2	A-E	Microbe Sequencing**	.fasta	<10MB	<100 files	<1GB
2	A-E	Bacterial Growth	.csv	<10MB	<500 files	<1GB
2	A-E	ELISA	.csv	<10MB	<500 files	<1GB
3	A-C	Protein Gel	.TIFF	<10MB	<200 files	<2GB
3	A-C	Western Blot (WB)	.TIFF	<10MB	<100 files	<1GB
3	A-C	Chromatography	.csv	<10MB	<100 files	<1GB
3	A	Heme Assay	.csv	<10MB	<100 files	<1GB
3	B	Analytical Ultracentrifugation	.csv	<10MB	<100 files	<1GB
3	C	Electron Microscopy	.TIFF	<2000MB	<40000 files	<80000 GB

*Approximate values. **Sequencing to confirm the sequence of plasmid or reference microbe strain.

The total size of the data collected is projected to be ≤125 TB (≤125000 GB).

AIM 1-3 - Data from Non-human Sources: This data includes all data generated from plasmids (e.g., sequencing), recombinant proteins (e.g., gels, WBs, ELISA, chromatography, electron microscopy), and reference microbial strains (e.g., sequencing, bacterial growth, gels, WBs, ELISA).

AIM 1D - ELISA Data from Human Biofluids: Only commercially available, de-identified human biofluids (e.g., Innovative Research Inc, Single Donor Human Synovial Fluid; Fisher Scientific, Fisher BioReagents, Human Serum - Normal Pool) will be used in this proposal. ELISA data in this proposal will not generate identifying information for any of these samples.

Scientific data that will be preserved and shared, and the rationale for doing so: *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

Except for the large Electron Microscopy data sets (4-6TB per data set, <100TB total), all other datasets described in **Table 1** will be preserved and shared through Digital Commons Data@Becker to enable validation of research results and accelerate future research directions. In addition to Digital

Commons Data@Becker, protein crystallography diffraction image data (TIFF) will be deposited into resources (Integrated Resource for Reproducibility in Macromolecular Crystallography [ProteinDiffraction.org], The Structural Biology Data Grid [data.sbgrid.org/]) that store diffraction image data related with Digital Object Identifiers (DOIs) related to PDB structures, while structure data (CIF) and models (mmCIF, pdb) will be deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). After processing Electron Microscopy data sets stored on Washington University (WU) Research Infrastructure Services (RIS) Active Storage, Electron Microscopy data sets will be temporarily moved to WU RIS Archive Storage before being made publicly available at the time of an associated publication or the end of the funding period, whichever comes first, through the Electron Microscopy Public Image Archive (EMPIAR, a public resource for raw Electron Microscopy images) and Electron Microscopy Data Bank (EMDB, a public repository for Electron Microscopy volumes and representative tomograms).

Metadata, other relevant data, and associated documentation: Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

For all datasets that will be deposited into Digital Commons Data@Becker, we will submit a README file to capture rich metadata and associated documents, such as data dictionaries and study protocols, along with datasets to facilitate the interpretation and reuse of the data.

The list of metadata that will be captured in a README file includes the following.

- General information: title, authors with ORCID iDs, organization (ROR), funder information (Funder Registry), award number, date of data collection, location of data collection, contextual description of the data
- Sharing/access information: licenses/restrictions placed on the data
- Data & file overview: list of file names and the relationship between files
- Methodological information: description of methods used for collecting/generation of data and processing the data, instrument and software-specific information to interpret the data, standards, and calibration information if appropriate, description of any quality-assurance procedures performed on the data, people involved with sample collection, processing, analysis and/or submission.
- Data-specific Information: number of variables, number of rows, variable list for each dataset

For protein crystallography data/models deposited into RCSB PDB, we will provide all metadata required for this data type by the repositories, such as information on date of collection, identities of the people who collected the data, location of data collection (e.g., synchrotron beamline, home source), protein identity (e.g., GenBank, UniProt identifiers), PDB identifier of solved structure (if deposited), detector type and serial numbers (S/N), image format, goniostat type, data-collection parameters (number of frames, oscillation-step size, goniostat orientation angles, 2θ offset, detector distance), resolution cutoffs, Rmerge, mean I, space group, program(s) used to refine the structure, R/Rfree, sample preparation data, crystallization conditions.

For electron microscopy data that will be deposited into EMPIAR and EMDB, we will provide all metadata required for this data type by the repositories, such as information on numbers of images, frames per image, image format, dimensions, pixel spacing, and type, and parameters essential for reprocessing, such as Cs, electron dose per frame, and gain-reference orientation.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

All data will be made available in open file formats (see **Table 1**), which can be accessed using publicly available software, and will not require the use of specialized tools to be accessed or manipulated. For example, fasta/CIF/mmCIF/pdb files can be opened using a text editor software such as Notepad and Microsoft Word, csv files can be opened using any spreadsheet program such as Google Sheets and Microsoft Excel, and TIFF files can be opened with an image file viewer such as Image J. Crystallographic Information File (CIF) uses standard text file format to store crystallographic structural data, while Macromolecular Crystallographic Information File (mmCIF) and Protein Data Bank (pdb) files use standard text file format to represent macromolecular structure data that describes the three-dimensional structures of molecules held in the Protein Data Bank (PDB).

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

We will use the community-accepted standards and ontologies in the fields that are available on FAIRsharing.org and convert all data files into open file formats to enable interoperability.

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository)

Electronic microscopy data will be shared in EMPIAR and EMDB, and protein crystallography will be shared in RCSB PDB. The rest of the data will be deposited in the Washington University School of Medicine (WUSM) institutional data repository Digital Commons Data@Becker, which is administered by the Bernard Becker Medical Library. Digital Commons Data@Becker is a generalist repository that accepts all types of data and shares the same platform as Mendeley Data, one of the GREI repositories supported by the NIH.

How scientific data will be findable and identifiable: Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

All above mentioned repositories provide searchabel study-level metadata for dataset discovery and assigns Digital Object Identifiers (DOIs) as persistent unique identifiers. Data will be discoverable

online through a standard web search of the study-level metadata as well as the persistent pointer from the DOI to the dataset.

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

All scientific data generated from this project will be made available as soon as possible, and no later than the time of an associated publication or the end of the funding period, whichever comes first. The duration of data preservation and sharing will be governed by the retention policies of the above-mentioned repositories we have selected. Currently, Digital Commons Data@Becker's retention policy ensures perpetual availability of all data shared in the repository, with contingency plans in place for potential changes in repository hosting.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

Data from Non-human Sources: There are no anticipated factors or limitations that will affect the access, distribution, or reuse of this scientific data generated by the proposal.

ELISA Data from Human Biofluids: Since all human biofluids used in this proposal will be obtained as de-identified samples from a commercial source, no human subjects' data will be available to be shared.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Controlled access will not be used. The data will be shared by unrestricted download with a Creative Commons license.

Protections for privacy, rights, and confidentiality of human research participants: If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

This research is not considered human subjects research and does not contain PII (Personally Identifiable Information). Therefore, no additional steps will be needed to ensure the confidentiality of human participants.

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Lead PI William Howard McCoy IV, M.D., Ph.D., ORCID: 0000-0001-5115-3793, will meet regularly with team members to ensure that data collection, management, and submission to the repositories follow FAIR data principles and occur in a manner compliant with this Data Management and Sharing Plan. Compliance will be evaluated annually during the award period, and progress toward the plan's DMS activities will be included in the annual Research Performance Progress Report (RPPR). In the event there are deviations from the approved DMS plan, we will seek re-approval of the revised DMS plan. At the project conclusion, the final progress report will summarize how the DMS objectives were fulfilled and provide links to the shared dataset(s).
