DMP du projet "Targeting CSC in breast cancer with 'Protein-Protein Interaction Inhibitor-like' compounds"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Renseignements sur le plan

Plan title	DMP du projet "Targeting CSC in breast cancer with 'Protein-Protein Interaction Inhibitor-like' compounds"
Fields of science and technology (from OECD classification)	
Language	eng
Creation date	2021-04-08
Last modification date	2023-01-20
Identifier	version 1 - 08/04/2021
License	Etalab Open License 2.0

Renseignements sur le projet

Project title	Targeting CSC in breast cancer with 'Protein-Protein Interaction Inhibitor-like' compounds
Acronym	bCSC targetting
Abstract	The "C. Ginestier and E. Charafe's" team at the CRCM (and other research groups) have confirmed that the aldehyde dehydrogenase (ALDH) enzymatic activity can be used to specifically identify breast CSC cells in the total tumor pool. These findings motivated the design of this research program, in strong collaboration with this team, aiming at identifying small molecule with anti-CSC properties in the context of breast cancer. Our goal will be to screen our library of chemical molecules (Fr-PPIChem) to identify compounds with cytotoxic effect on CSC (excluding anti-ALDH enzymatic activity). Inhibitors identified in the course of this research program will lead to i) useful chemical probes to help understanding CSC targeting, ii) innovative and efficient potential new drug candidates to be further developed, but also iii) identification of new protein/protein interaction with potential key roles in breast CSCs biology and breast cancer care.
Funding	• Institut National Du Cancer : Canceropole PACA - AAP Emergence 2022
Start date	2022-01-03
End date	2023-01-31
Partners	• Christophe Ginestier (201220152G) • Guillaume Pinna (201722499L)

Produits de recherche :

1. High Content Screening (Dataset)
2. Biochemistry/Enzymatic assay (Dataset)
3. Procedures for the chemical synthesis of small molecules (Text)
4. Proteomics (Jeu de données)
5. Tumorsphere Assay (Dataset)

Contributeurs

Name	Affiliation	Rôles
Betzi Stéphane - 0000-0001-5935-5058	CRCM	• DMP manager • Personne contact pour les données (Biochemistry, Tumorsphere Assay, Synthetic procedures, HCS, Proteomics) • Project coordinator

créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "Targeting CSC in breast cancer with 'Protein-Protein Interaction Inhibitor-like' compounds"

1. Data description and collection or re-use of existing data

High Content Screening

1a. How will new data be collected or produced and/or how will existing data be re-used?

Acquisition d'images avec un Microscope à épifluorescence automatisé (Operetta, Perkin Elmer)
Analyse d'images avec Harmony 3.1 (Perkin-Elmer), logiciel « tout en un » directement associé au microscope (pilotage du microscope + analyse d'images)
Harmony 3.1. pour la quantification des images (=data bruts)
Data bruts analysés sous excel (pré-analyse, contrôle qualité) et R 4.2.2 (analyse « massive » des données de crible).

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Format d'images brutes : Tiff organisés sous forme de base de données propriétaire (Harmony 3.1, Perkin Elmer).
Post-analyse d'images, export des quantifications au format CSV, séparateur tabulation.
Résultat final en format .xls ou .xlsx

Biochemistry/Enzymatic assay

1a. How will new data be collected or produced and/or how will existing data be re-used?

Biochemistry data mainly consists in images (gel pictures, crystal picture, microscope pictures) and text or table files of their analysis.
The enzymatic assay read from a Phearstar FR (BMG Labtech) will generate CSV table that will be analyzed using Graphpad PRISM.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Various images files *.png, *.jpg, *.tiff associated with text and table files as *.doc, *.txt, *.xlsx files. The total amount of data for one experiment is around 1Mo.
The enzymatic assay read from a Phearstar FR (BMG Labtech) will generate CSV table that will be analyzed using Graphpad PRISM to generate pzfx files that will be converted to CSV files for export.

Procedures for the chemical synthesis of small molecules

1a. How will new data be collected or produced and/or how will existing data be re-used?

Question sans réponse.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Question sans réponse.

Proteomics

1a. How will new data be collected or produced and/or how will existing data be re-used?

Mass spectra will be collected on a MALDI-TOF-TOF ULtraFlex III (Bruker), an ESI-LC-MS/MS Q Exactive Plus Orbitrap (Thermo Scientific), and an ESI-LC-MS/MS Fusion Lumos Tribrid Orbitrap (Thermo-Scientific).
Liquid chromatography coupled with mass spectrometry (HPLC-MS) will be done on Orbitrap Q-Exactive plus or Orbitrap Fusion Lumos Tribrib Mass Spectrometer from Thermo Fisher Scientific.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

The main software used to treat the collected data is proprietary (machine manifacture), and provides *.raw files. Treated data will be saved as *.txt, *.xml and *.mztab files. Final résultats will be exported as *.txt, *.xlsx, *.doc and images files (*.tiff, *.png, *.jpg).
The volume of proteomic data that will be generated during this project is estimated at several tens of Go.

Tumorsphere Assay

1a. How will new data be collected or produced and/or how will existing data be re-used?

Question sans réponse.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Question sans réponse.

2. Documentation and data quality

High Content Screening

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

The data will be accompanied by a “Readme.txt” file with the information on samples and experimental methodology to facilitate their re-use.

2b. What data quality control measures will be used?

Experiments will be performed according to "good laboratory practices" in line with the field standards. The consistency and quality of data will be documented in the data sets. Calibration experiments, repeated measurements, controls or peer review of data will be explained in the “Readme.txt” file.

Biochemistry/Enzymatic assay

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

The data will be accompanied by a “Readme.txt” file with the information on samples and experimental methodology to facilitate their re-use.

2b. What data quality control measures will be used?

Experiments will be performed according to "good laboratory practices" in line with the field standards. The consistency and quality of data will be documented in the data sets. Calibration experiments, repeated measurements, controls or peer review of data will be explained in the “Readme.txt” file.

Procedures for the chemical synthesis of small molecules

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Question sans réponse.

2b. What data quality control measures will be used?

Question sans réponse.

Proteomics

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

The data will be accompanied by a “Readme.txt” file with the information on samples and experimental methodology to facilitate their re-use.

2b. What data quality control measures will be used?

Experiments will be performed according to "good laboratory practices" in line with the field standards. The consistency and quality of data will be documented in the data sets. Calibration experiments, repeated measurements, controls or peer review of data will be explained in the “Readme.txt” file.

Tumorsphere Assay

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Question sans réponse.

2b. What data quality control measures will be used?

Question sans réponse.

3. Storage and backup during the research process

High Content Screening

3a. How will data and metadata be stored and backed up during the research?

The data are saved on the internal server of the platform. Backups are made daily.

3b. How will data security and protection of sensitive data be taken care during the research

NA

Biochemistry/Enzymatic assay

3a. How will data and metadata be stored and backed up during the research?

Data generated in the course of the project are documented on the INSERM Electronic Lab Notebook (ELN) called INSERM CLE by each participant on a daily basis to be shared within the consortium. The INSERM CLE is hosted safely in a SOC-compliant cloud data center on a dedicated INSERM server based in Montpellier under an agreement with the INSERM administration. All information is securely transferred (SSL), backed up daily and all entries and changes are recorded and time stamped automatically to ensure a full audit trail of your ELN. Everything entered and uploaded into the electronic lab notebook can be exported. We use this functionality to export locally the whole project files regularly (3 months basis).

3b. How will data security and protection of sensitive data be taken care during the research

Final data will be replicated on the team workstations as well as on the laboratory IT managed disk space (2x145 Tera octet configured as a replication node). This duplication on separate location and the replication on our NAS allows safety for the data. In addition, data generated in the course of the project are documented on the INSERM Electronic Lab Notebook (ELN) called INSERM CLE by each participant on a daily basis to be shared within the consortium. The INSERM CLE is hosted safely in a SOC-compliant cloud data center on a dedicated INSERM server based in Montpellier under an agreement with the INSERM administration. All information is securely transferred (SSL), backed up daily and all entries and changes are recorded and time stamped automatically to ensure a full audit trail of your ELN. Everything entered and uploaded into the electronic lab notebook can be exported. We use this functionality to export locally the whole project files regularly (3 months basis).

Procedures for the chemical synthesis of small molecules

3a. How will data and metadata be stored and backed up during the research?

Question sans réponse.

3b. How will data security and protection of sensitive data be taken care during the research

Question sans réponse.

Proteomics

3a. How will data and metadata be stored and backed up during the research?

The Mass spectra are saved on the internal server. Backups are made daily. The processed data are also stored at the institutional CNRS cloud service: https://mycore.core-cloud.net

3b. How will data security and protection of sensitive data be taken care during the research

The institutional cloud https://mycore.core-cloud.net meets requirements of data protection and safety defined by the organism (CNRS).

Tumorsphere Assay

3a. How will data and metadata be stored and backed up during the research?

Question sans réponse.

3b. How will data security and protection of sensitive data be taken care during the research

Question sans réponse.

4. Legal and ethical requirements, code of conduct

High Content Screening

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

NA

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The data belong to each of the two partners of the project: CRCM (CNRS, INSERM, AMU) and Institut de biologie François Jacob. They are research units/platform associated to several institutions (CNRS, INSERM, Aix-Marseille University, IPC for CRCM and

CEA for Institut de biologie François Jacob) and IP politics is defined in agreement with all institutional partners.
When appropriate, results will be submitted to institutional IP offices (Technology Transfert; SATT-Sud Est, INSERM Transfert, CNRS Innovation) for patent filing. These data won't be openly accessible due to confidential issues. Otherwise, data will be accessible on demand.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

The project is conducted in accordance with the charter of scientific integrity and research ethics established in partnership between CNRS, INSERM, CEA and universities.

Biochemistry/Enzymatic assay

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

NA

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The data belong to each of the three partners of the project: CRCM, CBS and MAP. They are research units/platform associated to several institutions (CNRS, INSERM, Aix-Marseille University, IPC) and IP politics is defined in agreement with all institutional partners.
When appropriate, results will be submitted to institutional IP offices (Technology Transfert; SATT-Sud Est, INSERM Transfert, CNRS Innovation) for patent filing. These data won't be openly accessible due to confidential issues. Otherwise, data will be accessible on demand.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

The project is conducted in accordance with the charter of scientific integrity and research ethics established in partnership between CNRS, INSERM and universities.

Procedures for the chemical synthesis of small molecules

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

Question sans réponse.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Question sans réponse.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

Question sans réponse.

Proteomics

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

NA

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The data belong to each of the three partners of the project: CRCM, CBS and MAP. They are research units/platform associated to several institutions (CNRS, INSERM, Aix-Marseille University, IPC) and IP politics is defined in agreement with all institutional partners.
When appropriate, results will be submitted to institutional IP offices (Technology Transfert; SATT-Sud Est, INSERM Transfert, CNRS Innovation) for patent filing. These data won't be openly accessible due to confidential issues. Otherwise, data will be accessible on demand.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

The project is conducted in accordance with the charter of scientific integrity and research ethics established in partnership between CNRS, INSERM and universities.

Tumorsphere Assay

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

Question sans réponse.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is

applicable?

Question sans réponse.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

Question sans réponse.

5. Data sharing and long-term preservation

High Content Screening

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

All data obtained will be preserved at the internal servers of each partner. The data supporting the results of all published articles will be made available through the journal publisher websites and on a long-term repository. After publication, raw and processed used in the publication/patent will be deposited whenever possible at the CINES or on another equivalent deposit that provide an associated DOI such as Zenodo.

5c. What methods or software tools are needed to access and use data?

Potential users will need the Harmony 3.1 (PerkinElmer) software to read raw data as organized in the proprietary database, however this set of data consist of collection of standard TIFF files.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

After publication, useful raw and processed data used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo.

Biochemistry/Enzymatic assay

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

All data obtained will be preserved at the internal servers of each partner. The data supporting the results of all published articles will be made available through the journal publisher websites and on a long-term repository. After publication, raw and processed used in the publication/patent will be deposited whenever possible at the CINES or on another equivalent deposit that provide an associated DOI such as Zenodo.

5c. What methods or software tools are needed to access and use data?

Most of the biochemistry experiments will use free format files as outputs such as *.txt, *.xlsx, *.doc and images files (*.tiff, *.png, *.jpg).

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo.

Procedures for the chemical synthesis of small molecules

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Question sans réponse.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Question sans réponse.

5c. What methods or software tools are needed to access and use data?

Question sans réponse.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Question sans réponse.

Proteomics

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo. Proteomics data will be specifically accessible via the Consortium PRIDE PRoteomics IDEntifications database (www.proteomexchange.org)

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo. Proteomics data will be specifically accessible via the Consortium PRIDE PRoteomics IDEntifications database (www.proteomexchange.org)

5c. What methods or software tools are needed to access and use data?

Although the need for a proprietary tool is required to open the direct raw data from the machine, free format will be used to generate output files for storage such as *.txt, *.xlsx, *.doc and images files (*.tiff, *.png, *.jpg).

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

After publication, useful raw and processed used in the publication/patent will be deposited at the CINES or on another equivalent deposit that provides an associated DOI such as Zenodo. Proteomics data will be specifically accessible via the Consortium PRIDE PRoteomics IDEntifications database (www.proteomexchange.org)

Tumorsphere Assay

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Question sans réponse.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Question sans réponse.

5c. What methods or software tools are needed to access and use data?

Question sans réponse.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Question sans réponse.

6. Data management responsibilities and resources

High Content Screening

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Partner 1 (CRCM) and Partner 2 (Institut de biologie François Jacob) share the responsibility for the management and sharing of raw and processed data. The co-ordination of data management responsibilities across both partners is assured by the coordinator of the project.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The partners will devote a part of their research time to prepare the data for sharing/preservation and publishing them on public repositories. The data management is taken care locally by each team laboratory IT support, therefore this process does not impact the project funding nor require local management ressource.

Biochemistry/Enzymatic assay

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Biochemistry raw and processed data are managed by Stephane Betzi (CNRS researcher) for the CRCM.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The partners will devote a part of their research time to prepare the data for sharing/preservation and publishing them on public repositories. The data management is taken care locally by each team laboratory IT support, therefore this process does not impact the project funding nor require local management ressource.

Procedures for the chemical synthesis of small molecules

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Question sans réponse.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Question sans réponse.

Proteomics

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Partner 1 (CRCM), Partner 2 (CBS) and Partner 3 (MAP) share the responsibility for the management and sharing of raw data and spectra images. The co-ordination of data management responsibilities across both partnets is assured by the coordinator of the project.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be

FAIR (Findable, Accessible, Interoperable, Re-usable)?

The partners will devote a part of their research time to prepare the data for sharing/preservation and publishing them on public repositories. The data management is taken care locally by each team laboratory IT support, therefore this process does not impact the project funding nor require local management ressource.

Tumorsphere Assay

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Question sans réponse.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Question sans réponse.