
LipInTB

Plan de gestion de données créé à l'aide de DMP OPIDoR

Créateur du PGD : Jean-François Cavalier

Affiliation du créateur principal : CNRS

Modèle du PGD : ANR - Modèle de PGD (français)

Dernière modification du PGD : 17/12/2019

Financier : Agence nationale de la recherche (ANR)

Numéro de subvention : ANR-19-CE44-0011

Résumé du projet :

La tuberculose causée par la bactérie pathogène *Mycobacterium tuberculosis* (*M. tb*) est la maladie infectieuse la plus meurtrière au monde. L'infection par *M. tb* conduit le plus souvent à la formation de granulomes dans les poumons, où certains macrophages infectés accumulent des lipides dans des corps lipidiques (LB) formant des cellules spumeuses. Au sein de ces macrophages spumeux (FM), les bacilles peuvent persister dans un état non-réplicatif pendant des décennies avant de se réactiver et causer la maladie. Une meilleure compréhension de la persistance des bacilles à l'intérieur de ces FM est donc nécessaire pour lutter contre la tuberculose.

Pour persister dans les FM, *M. tb* hydrolyse les lipides de l'hôte en acides gras, qu'elle réabsorbe et stocke sous forme d'inclusions lipidiques intracytoplasmiques (ILI). Ces réserves lipidiques pourront ensuite être réutilisées lors de la phase de réactivation. Des résultats récents ont suggéré un lien direct entre la présence de ces ILI et l'incapacité des mycobactéries à se diviser. Ce dernier point est d'une importance capitale pour la persistance des mycobactéries au sein du granulome. Le contrôle de l'hydrolyse des lipides de l'hôte ainsi que celle des ILI au cours de l'infection commence à être bien documentée et serait assuré par des enzymes lipolytiques mycobactériennes. Depuis une vingtaine d'années, ces enzymes lipolytiques, responsables de la libération d'acides gras à longues chaînes, font l'objet d'intenses recherches. Elles seraient impliquées dans la relation hôte-pathogène et joueraient un rôle majeur dans la physiopathologie de la maladie lors des phases active et persistante de l'infection, faisant de ces enzymes de véritables cibles thérapeutiques. Cependant les mécanismes moléculaires qui régissent tous ces processus restent inconnus. Par conséquent, trouver des moyens pour bloquer l'action de ces enzymes ouvrirait la voie vers la découverte de nouveaux traitements contre la tuberculose.

Dans ce contexte, nous avons récemment découvert deux nouvelles familles de molécules non cytotoxiques pour les cellules hôtes, et qui ciblent diverses enzymes mycobactériennes impliquées dans le métabolisme des lipides avec des activités antituberculeuses très prometteuses.

Notre projet, impliquant 5 partenaires, vise à utiliser ces inhibiteurs afin i) d'identifier et de valider *in vitro* et *in vivo* les enzymes impactées par nos composés et conduisant à la mort de *M. tb* en extracellulaire et dans des macrophages infectés; et ii) de les utiliser comme sondes pour étudier le métabolisme lipidique de

M. tb lors des phases de réplication active, de latence et de réactivation dans les macrophages infectés.

Des données préliminaires laissent supposer que ces inhibiteurs ont des effets importants sur la maladie en modifiant profondément la réplication des mycobactéries dans les cellules hôtes, leur entrée en phase de persistance et/ou la réactivation des bacilles dans les FM qui sont des enjeux majeurs pour comprendre la vulnérabilité et l'évolution générale de la tuberculose.

Les résultats attendus sont en lien direct avec la pathogénie des bacilles pendant les diverses phases de l'infection. De plus, la parfaite complémentarité des équipes sera un atout majeur pour la réalisation de ce projet.

Enfin, les connaissances issues de ce projet pourront conduire au développement de nouvelles stratégies thérapeutiques pour éliminer les bacilles en réplication active ou en latence chez des individus infectés.

Chercheur Principal : Jean-François Cavalier

Contact pour les Données : Jean-François Cavalier

Droits d'auteur

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

1. Description des données et collecte ou réutilisation de données existantes

Les données recueillies dans le cadre de l'étude proposée comprendront la synthèse de composés chimiques (protocoles et caractérisation) et leur utilisation en tant que sonde moléculaire. Chaque composé synthétisé sera purifié et entièrement caractérisé pour confirmer leur structure et leur pureté chimique à l'aide de méthodes spectroscopiques et chromatographiques. Les données spectroscopiques typiques comprennent la RMN 1H, la RMN 13C, la spectrométrie IR et la spectrométrie de masse ; les techniques chromatographiques comprennent la chromatographie en phase liquide couplée à la spectrométrie de masse (HPLC-MS).

Les composés seront ensuite utilisés en tant que sonde moléculaire pour étudier le métabolisme des lipides chez *M. tb*. Les données préexistantes et déjà publiées seront également utilisées pour la validation des protéines cibles de nos composés. Une fois les protéines validées, elles seront cristallisées et leurs structures 3-D respectives seront déterminées par la méthode de diffraction aux rayons-X.

Les données seront collectées par chaque scientifique impliqué dans le projet à l'aide de logiciel adapté à chaque instrument. Les protocoles expérimentaux sont conservés sous forme papier (dans les cahiers de laboratoire) et version électronique (.docx, .pdf). Les données RMN seront stockées sous forme électronique (fichiers FID bruts et format PDF) et en laboratoire sous forme de copies papier des spectres. Les fichiers FID bruts collectés sur les appareils Bruker peuvent être traités par n'importe quel logiciel de RMN, et seront disponibles sur demande. Les fichiers PDF contiendront à la fois des images des spectres originaux et des tableaux numériques détaillant le déplacement chimique, les constantes de couplage, les valeurs de masse et autres interprétations numériques des données.

Les données biochimiques générées seront : des données expérimentales (formats csv, xls), des données textuelles (format pdf, docx, pptx), des données de protéomique, des données cristallographiques, des images (microscopie confocale et électronique - format png). Les données protéomiques seront de plusieurs types, les fichiers bruts en sortie des spectromètres de masse (.raw), les données traitées (.txt, .xml, .mztab) et les résultats (.txt, .xlsx, .docx et images). Le volume des données protéomiques qui seront générés pendant ce projet est estimé à plusieurs Go.

Les données de diffraction aux rayons X seront stockées sous forme électronique (format .cbf et .h5) à la fois par les sites de rayonnement synchrotron (Soleil et ESRF), et au sein du laboratoire du partenaires du projet. Le volume des données de diffraction qui seront générés pendant ce projet est estimé à plusieurs To. Selon les directives H2020 les sites de rayonnement synchrotron doivent donner accès sur demande aux données de diffraction trois ans après leur acquisition. Les fichiers avec les coordonnées atomique des structure 3-D des protéines seront ensuite déposés à la Protein Data Bank (<https://www.rcsb.org/>) au format PDB.

Dans la mesure du possible des formats standards et ouverts seront privilégiés à des fins de partage et de réutilisation.

2. Documentation et qualité des données

Cf. réponse 1b

Au niveau des données de RMN, de spectrométrie de masse et de diffraction aux rayons X, des métadonnées comme les paramètres d'acquisition, le type d'appareillage, la version des logiciels seront incluses dans les fichiers bruts. Pour ce qui concerne les données de diffraction, les métadonnées seront également accessibles *via* le système de gestion des informations de laboratoire ISPYB (<https://www.esrf.eu/ispyb>). Toutes les métadonnées concernant l'obtention de la structure 3-D d'une protéine seront incluses dans les fichiers PDB.

Les Bonnes Pratiques de Laboratoire seront suivies pour le contrôle et la qualité des données.

Au niveau des données de spectrométrie de masse, la qualité et la conformité des données seront contrôlées par les divers contrôles qualités et les calibrations des appareils réalisés régulièrement. De plus des répliques biologiques et techniques seront réalisés pour des études de protéomique quantitative.

Au niveau de la synthèse de nouvelles molécules, les spectromètres RMN sont calibrés et révisés par le Laboratoire de Mesure Physique de l'IBMM.

Concernant les mesures de diffraction aux rayons X, les lignes de lumières sont calibrées au quotidien par les

responsables de ligne. Le traitement des données à la volée, fournissant un grand nombre d'indicateurs de qualité, permet également de contrôler la qualité des données.

3. Stockage et sauvegarde pendant le processus de recherche

Pour la durée du projet ANR les données générées seront hébergées sur les postes de travail chiffrés des membres du projet LipInTB, et pour les données de protéomique et de cristallographie sur des serveurs dédiés et sécurisés. Des sauvegardes sur un serveur propre à chaque unité seront effectuées à intervalles réguliers. Les données de diffraction aux rayons X seront également sauvegardées par les sites de rayonnement synchrotron.

Les données seront stockées sur les postes de travail chiffrés de chaque responsable scientifique du projet, ou sur des serveurs dédiés et sécurisés pour les données de protéomique et de cristallographie.

4. Exigences légales et éthiques, codes de conduite

Non applicable.

Chaque partenaire respectera les règles de sa tutelle scientifique respective, en matière de propriété intellectuelle et de brevet (INSERM, CNRS ET AMU).

Tous les partenaires seront associés aux publications en fonction de leurs contributions respectives. Afin de promouvoir l'accessibilité, des publications en libre accès seront privilégiées, dans la mesure du possible. A l'issue du processus de publications et en respectant les politiques éditoriales des éditeurs, les publications issues du projet seront déposées dans une archive ouverte, libre d'accès (HAL, ...)

Non applicable.

5. Partage des données et conservation à long terme

Les résultats obtenus seront publiés dans des revues à comité de lecture appropriées. A l'issue du processus de publications et en respectant les politiques éditoriales des éditeurs, les publications issues du projet seront déposées dans une archive ouverte, libre d'accès (HAL, ...)

Toutes les structures protéiques seront librement disponibles car déposées dans RCSB Protein Data Bank (<https://www.rcsb.org/>).

Les données de protéomique seront également accessibles *via* le Consortium PRIDE : PRoteomics IDentifications database (www.proteomexchange.org).

L'ensemble des données générées seront uniques et devront donc être conservées sans mention de délai.

Les données cristallographiques et protéomiques seront conservées sur des entrepôts de données accessibles à tous (PRIDE-EBI pour la protéomique et RCSB Protein Data Bank pour la cristallographie).

Dans la mesure du possible des formats standards et ouverts seront privilégiés à des fins de partage et de réutilisation.

Non applicable.

6. Responsabilités et ressources en matière de gestion des données

Chaque responsable scientifique sera en charge de l'archivage des données le concernant.

Chaque responsable scientifique sera en charge de la gestion des données le concernant.