

[CDD] Ingénieur·e des données de spectrométrie de masse pour l'élucidation structurale

Contexte : La création de bases de données de spectrométrie de masse en tandem (MS/MS) est un enjeu majeur pour l'entraînement de modèles d'IA afin de prédire la structure des petites molécules [1]. La constitution de telles bases de données nécessite le développement de workflows computationnels pour planifier les acquisitions, traiter les données, et les formater [2]. Dans le cadre de l'infrastructure de métabolomique (MetaboHUB) et du projet d'IA pour l'élucidation structurale (e-Lucid), de nouvelles données MS/MS multi-modales sont en cours d'acquisition (multiplicité des technologies et des conditions de fragmentation).

Sujet : L'ingénieur·e sera chargé·e de développer les workflows computationnels haut-débit pour traiter ces données, créer les bases de données et les diffuser dans la communauté. Les workflows incluront notamment les modules de planification des expériences, de recherche des métadonnées associées aux molécules, de traitement des spectres, de leur matching avec les bases de données spectrales publiques, de leur indexation, de leur harmonisation, de leur formatage pour l'IA au sein de bases de données compatibles avec l'entraînement des modèles, et de la diffusion de ces bases de données et des bibliothèques logicielles de traitement dans un contexte FAIR auprès des communautés informatiques et expérimentales.

Références :

[1] Rutz, A. et al. (2026) MultiMS2: A curated multi-modal, multi-energy spectral library for metabolomics. *Gigascience*. <https://doi.org/10.1093/gigascience/giag069>.

[2] Brungs, C. et al. (2025) MSnLib: efficient generation of open multi-stage fragmentation mass spectral libraries. *Nat Methods*, 1–4. <https://doi.org/10.26434/chemrxiv-2024-l1tqh-v3>.

Profil : Nous recherchons un·e ingénieur·e bioinformaticien·ne avec une solide formation en programmation Python, traitement des données, gestion de workflow (e.g., Nextflow, Galaxy) et bases de données (e.g., PostgreSQL), et si possible une expérience des données de spectrométrie de masse MS/MS.

Laboratoire d'accueil : L'équipe [Odiscé](#) du [CEA](#) Saclay, membre de l'infrastructure nationale de métabolomique [MetaboHUB](#), est spécialisée depuis plus de 15 ans dans le développement de méthodes en informatique mathématique (traitement du signal, machine learning, chémoinformatique, gestion de workflows) et de bibliothèques logicielles innovantes en sciences des données de spectrométrie de masse pour la métabolomique de population. Les données et algorithmes sont mis à disposition de la communauté notamment grâce à une [suite logicielle](#), également disponible sous [Galaxy/Workflow4Metabolomics](#).

Durée : 24 mois

Date de début du contrat : janvier 2027

Date limite de candidature : 30/10/2026

Contact : Merci d'adresser votre candidature (CV détaillé avec les expériences professionnelles, lettre de motivation, coordonnées des personnes référentes) à :

Etienne Thévenot (etienne.thevenot@cea.fr)

Equipe sciences des données de spectrométrie de masse pour la métabolomique (Odiscé)

Laboratoire Innovations en spectrométrie de Masse pour la Santé (LI-MS)

CEA, Centre de Saclay, F-91191 Gif sur Yvette, France

Team: <https://odisce.github.io/>

GitHub: <https://github.com/odisce>