

Sujet Master/TFE 2023-2024

Machine-learning-based analysis of Pseudo Mass-Spectrum Images for Targeted Peptides Identification

Encadrants :

- Stéphane Derrode (stephane.derrode@ec-lyon.fr), École Centrale de Lyon, laboratoire LIRIS (CNRS UMR 5205, <https://liris.cnrs.fr>).
- Jérôme Lemoine (jerome.lemoine@univ-lyon1.fr), Université Claude Bernard Lyon 1, Institut des Sciences Analytiques (ISA, CNRS UMR 5280, <https://isa-lyon.fr>).

Date butoir pour candidater : vendredi 15 décembre 2023.

Contexte

Ce projet est plus particulièrement lié à un programme de recherche interdisciplinaire (financement ANR, RHU IDBIORIV - 2019-2026) incluant des médecins, des microbiologistes cliniciens et des statisticiens (Institut des Agents Infectieux, Centre International de Recherche en Infectiologie), et dont l'objectif principal est de développer de nouvelles méthodes basées sur la spectrométrie de masse et des pipelines de traitements automatiques pour réduire le délai de l'étape de diagnostic des infections du sang, des urines ou toute autre matrice biologique.

Le mode d'acquisition SWATH/DIA permet d'identifier des peptides rapporteurs sans recourir à la construction de banques spectrales. Le projet de recherche vise au développement d'algorithmes, basé sur l'apprentissage profond (*deep learning*), permettant de traiter les données brutes en mode DIA, et capable de fournir des informations directement interprétables par les médecins infectiologues. Il s'agira en particulier d'identifier des micro-organismes pathogènes et prédire le niveau des résistances aux antibiotiques de ces mêmes pathogènes.

Sujet du stage

Les données issues d'un chromatogramme reconstitué sur la valeur m/z des ions se présentent sous la forme de séries temporelles --représentant le temps de rétention (min.) pour différents rapports masses-sur-charge (m/z). Les méthodes d'analyse traditionnelles consistent en 4 étapes : (1) prétraitement des données brutes, (2) nettoyage des données, (3) identification des peptides, notamment par *peak-picking* [6], et (4) diagnostic. L'identification automatique des peptides est la tâche la plus ambitieuse, du fait de la complexité du bruit et des interférences liées à la complexité de la matrice biologique [4] !

Étant donné les volumes des données manipulées, ainsi que les bruits et artefacts spécifiques d'acquisition d'un chromatogramme (p. ex. dérive des temps de rétention des signaux), des progrès sont nécessaires pour reconnaître automatiquement les peptides cibles [3]. Dans ce projet, nous proposons une approche différente de la méthode traditionnelle, qui consiste à transformer les gros volumes de données 1D en images monovaluées sur lesquelles des méthodes de reconnaissance de formes et d'apprentissage statistique et/ou profond pourront s'appliquer (cf. Figure 1).

Les résultats seront objectivés par le calcul de performances quant au de taux d'identification des peptides cibles (sensibilité et spécificité), en bout de chaîne, grâce à la connaissance précise des échantillons biologiques analysés au sein des équipes de l'Institut des Sciences Analytiques (J. Lemoine) et de l'Institut des Agents Infectieux de l'Hôpital de la Croix Rousse (F. Vandenesch).

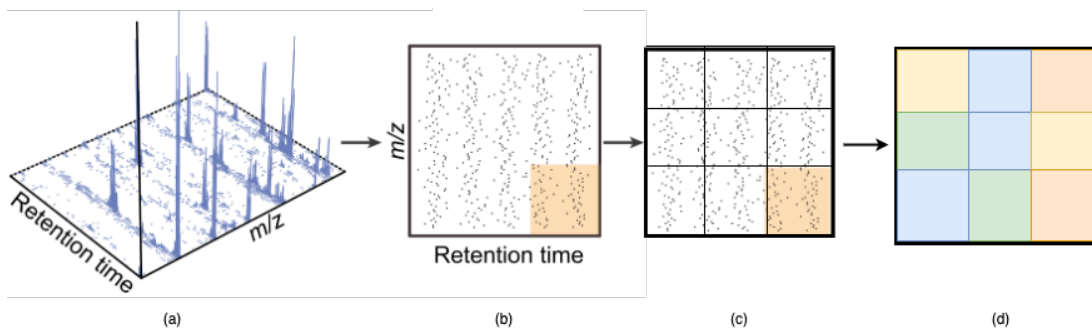


Figure 1: Processus de transformation des données originales en image.

Compétences requises

L'étudiant.e devra disposer de bonnes compétences dans les domaines du Machine Learning. Une expérience certaine de la programmation en langage Python est également requise pour implémenter les solutions envisagées (*pytorch / tensorflow*), les évaluer et les comparer sur le corpus de données.

Informations pratiques :

- Lieu du stage : Laboratoire LIRIS - Ecole Centrale de Lyon (Ecully). Des déplacements au laboratoire ISA (site de La Doua) sont également prévus.
- Période de stage : Février-Mars 2024, pour une durée de 5 à 6 mois.
- Rémunéré : Oui (4.05 euros/h, 35h/semaine, soit environ 580euros /mois)

Pour candidater :

- Merci d'envoyer votre CV aux DEUX (2) encadrants (adresses mail ci-dessus).
- **Date butoir pour candidater : vendredi 15 décembre 2023.**
- Les étudiants présélectionnés seront invités à un entretien, et classés selon leur ordre de mérite et leurs motivations pour le sujet.
- L'étudiant.e qui s'engagera sur le sujet devra suivre une procédure administrative pour intégrer le LIRIS.

Éléments bibliographiques :

- [1] Shen X, Shao W, Wang C, Liang L, Chen S, Zhang S, Rusu M, Snyder MP. *Deep learning-based pseudo-mass spectrometry imaging analysis for precision medicine*. Brief Bioinform. 2022 Sep 20;23(5):bbac331.
- [2] Wishart, David S., *Emerging applications of metabolomics in drug discovery and precision medicine*, Nature reviews Drug discovery 15.7 (2016): 473-484.
- [3] Alseekh, S., Aharoni, A., Brotman, Y. et al., *Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices*, Nat Methods 18, 747–756 (2021).
- [4] Chaleckis, Romanas, et al., *Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics*, Current opinion in biotechnology, 55 (2019): 44-50.
- [5] Wang, H, Yandong Y, and Zheng-Jiang Z. *Encoding LC–MS-Based Untargeted Metabolomics Data into Images toward AI-Based Clinical Diagnosis*, Analytical Chemistry (2023).
- [6] Guo, J, et Tao H. *Mechanistic Understanding of the Discrepancies between Common Peak Picking Algorithms in Liquid Chromatography–Mass Spectrometry-Based Metabolomics*, Analytical Chemistry 95, no 14 (2023): 5894-5902.