

# Traitement convergent des données de RMN et LC-MS pour la métabolomique

## Thèse en Chemo-informatique à Nantes

### Contexte Universitaire

Dans le cadre d'une collaboration entre le laboratoire laboratoire Mer, Molécules, Santé (MMS – EA2160) équipe ChiChaMVa (<https://mms.univ-nantes.fr/equipe-4-chichamva/>) et le laboratoire CEISAM (UMR UMR CNRS 6230) équipe MIMM (<https://ceisam.univ-nantes.fr/en/mimm-team/>), nous recherchons pour cette thèse **soumise au concours** de l'école doctorale SML (<https://ed-sml.doctorat-bretagne.fr/>) un candidat spécialisé souhaitant effectuer une thèse en chemo-informatique, bio-informatique ou métabolomique.

### Contexte thématique

La métabolomique est une approche récente pour la compréhension des systèmes biologiques, visant à identifier et quantifier les métabolites (petites molécules impliquées dans le métabolisme) dans des échantillons complexes (biofluides, extraits de cellules, de plantes ou de produits naturels...). Les workflow d'analyse en métabolomique reposent sur des techniques analytiques de pointe : la résonance magnétique nucléaire (RMN) et la chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS). Malgré d'importants progrès dans l'intégration des données issues de ces deux techniques, le défi majeur reste le passage à l'annotation des biomarqueurs mises en évidence au sein de mélanges complexes [1-3]. Un grand nombre d'approches sont actuellement mise en place du côté de l'analyse de données LC-HRMS, profitant en partie de l'unicité de structure des données entre différentes méthodes d'analyse [2, 4, 5]. Du côté de la RMN, les multiples approches complémentaires permettant l'analyse monodimensionnelle (proton, carbon, etc.) ou bidimensionnelle (proton-proton, carbon-proton – HSQC, HMBC, etc.) posent d'autres problèmes de fusion des données d'un point de vue spectrale (différence de résolution, de recouvrement, etc.) [6-8].

Cette thèse se propose de développer de nouvelles approches permettant de combiner les données RMN et LC-HRMS en utilisant une approche basée sur les informations structurales fournies par les stratégies d'analyses RMN. Pour cela, le travail se focalisera en particulier sur les stratégies de fusion des données RMN (incluant de nouvelles approches de RMN 2D rapide [9, 10]) pour permettre l'isolement de données structurales (système de spin) via la création de réseaux de spin RMN [11] pouvant se rapprocher par exemple des approches MS2LDA utilisées en LC-HRMS [12]. Ces stratégies permettront ensuite la fusion des données RMN et LC-HRMS suivant une approche basée sur la structure des molécules chimiques présentes dans des mélanges complexes ; à opposer aux stratégies basées sur des approches statistique de fusion de données [13-16].

Ces approches seront mises en place dans le contexte de l'étude du métabolisme spécialisé de champignons d'origine marine afin de caractériser au mieux la diversité chimiques produites par ces champignons [17, 18]. Il s'agit ici d'une étape clé préalable à la valorisation des molécules naturelles

[2, 19-21]. Dans ce cadre, un grand nombre de jeu de données ont d'ores-et-déjà été acquis et serviront de base au développement des nouvelles approches d'analyse de données RMN et LC-HRMS. Actuellement de fortes attentes existent dans ce domaine aussi bien coté RMN que LC-HRMS. Ainsi, les méthodes développées devraient présenter un fort intérêt pour une vaste communauté de chercheurs.

## Profil recherché

Nous recherchons un étudiant en Master2 Bio-Informatique, Chemo-informatique et ou métabolomique, motivé, créatif, force de proposition pour travailler sur le traitement de données spectrales multi-échelle dans un environnement multidisciplinaire. Des compétences en programmation sur R, Python et/ou Matlab sont indispensables pour cette thèse qui se déroulera essentiellement sur l'analyse de ces données.

La thèse sera effectuée au sein du laboratoire MMS à Nantes en collaboration étroite avec le laboratoire CEIAM, et sera réalisé dans le contexte de collaboration inter-plateau de la plateforme de métabolomique CORSAIRE (<https://www.pf-corsaire.org/>) de BiogenOuest (<https://www.biogenouest.org/>). Ces deux laboratoires sont fortement intégrés au réseau francophone de métabolomique et fluxomique (RFMF), offrant ainsi une vaste communauté d'utilisateurs potentiels pour les méthodes qui seront développées.

**Cette thèse est mise au concours de l'école doctorale Sciences de la Mer et du Littoral, et le succès au concours permettra d'initier celle-ci qui débutera en septembre/octobre 2021 pour 3 ans.**

## Contact :

**Directeur :** Dr. Samuel BERTRAND (HDR), MMS, Université de Nantes, [samuel.bertrand@univ-nantes.fr](mailto:samuel.bertrand@univ-nantes.fr), +33(0)2 51 12 56 89

**Co-directeur :** Prof. Patrick Giraudeau, CEISAM, Université de Nantes, [patrick.giraudeau@univ-nantes.fr](mailto:patrick.giraudeau@univ-nantes.fr), +33(0)2 51 12 57 09

## Bibliographie

1. Wolfender, J.-L., et al., *Current approaches and challenges for the metabolite profiling of complex natural extracts*. Journal of Chromatography A, 2015. **1382**: p. 136–164.
2. Wolfender, J.-L., et al., *Accelerating metabolite identification in natural product research: toward an ideal combination of LC-HRMS/MS and NMR profiling, in silico databases and chemometrics*. Analytical Chemistry, 2019. **91**(1): p. 704-742.
3. Letertre, M.P.M., G. Dervilly, and P. Giraudeau, *Combined Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry Approaches for Metabolomics*. Analytical Chemistry, 2020.
4. Aron, A.T., et al., *Reproducible molecular networking of untargeted mass spectrometry data using GNPS*. Nature Protocols, 2020. **15**(6): p. 1954-1991.
5. Huber, F., et al., *Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships*. PLOS Computational Biology, 2021. **17**(2): p. e1008724.
6. Willoughby, P.H., M.J. Jansma, and T.R. Hoyer, *A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts*. Nature Protocols, 2014. **9**(3): p. 643-660.
7. Halabalaki, M., et al., *Recent advances and new strategies in the NMR-based identification of natural products*. Current Opinion in Biotechnology, 2014. **25**(0): p. 1-7.
8. Garcia-Perez, I., et al., *Identifying unknown metabolites using NMR-based metabolic profiling techniques*. Nature Protocols, 2020.

9. Marchand, J., et al., *Multidimensional NMR approaches towards highly resolved, sensitive and high-throughput quantitative metabolomics*. *Current Opinion in Biotechnology*, 2017. **43**: p. 49-55.
10. Martineau, E., J.-N. Dumez, and P. Giraudeau, *Fast quantitative 2D NMR for metabolomics and lipidomics: A tutorial*. *Magnetic Resonance in Chemistry*, 2020. **58**(5): p. 390-403.
11. Egan, J.M., et al., *Development of an NMR-Based Platform for the Direct Structural Annotation of Complex Natural Products Mixtures*. *Journal of Natural Products*, 2021.
12. van der Hooft, J.J.J., et al., *Topic modeling for untargeted substructure exploration in metabolomics*. *Proceedings of the National Academy of Sciences*, 2016. **113**(48): p. 13738-13743.
13. Bertrand, S., et al., *Statistical correlations between HPLC activity-based profiling results and NMR/MS microfractions data to deconvolute bioactive compounds in mixture*. *Molecules*, 2016. **21**(3): p. 259.
14. Chen, H.W., et al., *Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation*. *Rapid Communications in Mass Spectrometry*, 2006. **20**(10): p. 1577-1584.
15. Boccard, J. and D.N. Rutledge, *A consensus OPLS-DA strategy for multiblock Omics data fusion*. *Analytica Chimica Acta*, 2013. **769**: p. 30-39.
16. Marti, G., et al., *Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance*. *Food Chemistry*, 2014. **150**: p. 235-245.
17. Sigwart, J.D., et al., *Unlocking the potential of marine biodiscovery*. *Natural Product Reports*, 2021.
18. Carroll, A.R., et al., *Marine natural products*. *Natural Product Reports*, 2021.
19. Lianza, M., et al., *The Three Pillars of Natural Product Dereplication. Alkaloids from the Bulbs of *Urceolina peruviana* (C. Presl) J.F. Macbr. as a Preliminary Test Case*. *Molecules*, 2021. **26**(3): p. 637.
20. Atanasov, A.G., et al., *Natural products in drug discovery: advances and opportunities*. *Nature Reviews Drug Discovery*, 2021.
21. Hubert, J., J.-M. Nuzillard, and J.-H. Renault, *Dereplication strategies in natural product research: How many tools and methodologies behind the same concept?* *Phytochemistry Reviews*, 2015. **16**(1): p. 55-95.