

Convergent processing of NMR and LC-MS data for metabolomics

PhD in Chemo-informatics in Nantes

Context

This PhD project is part of a collaboration between the Mer, Molécules, Santé laboratory (MMS - EA2160), ChiChaMVa team (<https://mms.univ-nantes.fr/equipe-4-chichamva/>) and the CEISAM laboratory (UMR UMR CNRS 6230), MIMM team (<https://ceisam.univ-nantes.fr/en/mimm-team/>). We are looking for a PhD candidate to compete for the SML doctoral school scholarship (<https://ed-sml.doctorat-bretagneoire.fr/>), which wishing to perform a thesis in chemo-informatics, bioinformatics and metabolomics.

Project

Metabolomics is a recent approach aiming to understand biological systems, by identify and quantify metabolites (small molecules involved in metabolism) in complex samples (biofluids, extracts from cells, plants or natural products, etc.). The metabolomic workflows are based on cutting-edge analytical techniques: nuclear magnetic resonance (NMR) and liquid chromatography coupled with high resolution mass spectrometry (LC-HRMS). Despite significant progress in integrating data from these two techniques, the major challenge remains the annotation of biomarkers identified within complex mixtures [1-3]. A large number of approaches are currently being implemented on the LC-HRMS data analysis side, in part benefiting from the uniqueness of the data structure between different analysis methods [2, 4, 5]. Regarding NMR, the multiple complementary approaches allowing one-dimensional (proton, carbon, etc.) or two-dimensional (proton-proton, carbon-proton - HSQC, HMBC, etc.) analysis pose other data fusion problems from a spectral point of view (difference in resolution, overlap, etc.) [6-8].

This thesis aims to develop new approaches allowing to combine NMR and LC-HRMS data using an approach based on structural information provided by NMR analysis strategies. For this, the work will focus in particular on NMR data fusion strategies (including new fast 2D NMR approaches [9, 10]) to allow the isolation of structural data (spin system) via the creation of NMR spin networks [11]. Such approaches may be similar, for example, to the MS2LDA approaches used in LC-HRMS [12]. These strategies will then allow the fusion of the NMR and LC-HRMS data following an approach based on the structure of the chemical molecules present in complex mixtures; to be opposed to strategies based on statistical approaches to data fusion [13-16].

These approaches will be implemented in the context of the study of the specialized metabolism of marine fungi to better characterize the chemical diversity produced by these fungi [17, 18]. This is a key step prior to valuing natural molecules [2, 19-21]. In this context, a large number of datasets have already been acquired and will serve as a basis for the development of new approaches to NMR and LC-HRMS data analysis. Currently, high expectations exist in this field, both NMR and LC-HRMS rated. Thus, the methods developed should be of great interest to a large community of researchers.

Requirement

We are looking for a Master student in Bio-Informatics, Chemo-informatics and/or metabolomics, motivated, creative, force of proposal to work on the processing of multiscale spectral data in a multidisciplinary environment. Programming skills in R, Python and/or Matlab are essential for this thesis, which will focus on the analysis of this data.

The thesis will be carried out in the MMS laboratory in Nantes in close collaboration with the CEIAM laboratory, and will be carried out in the context of inter-platform collaboration of the CORSAIRE metabolomics platform (<https://www.pf-corsaire.org/>) from BiogenOuest (<https://www.biogenouest.org/>). These two laboratories are strongly integrated into the French-speaking network of metabolomics and fluxomics (RFMF), thus offering a large community of potential users for the methods that will be developed.

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

This thesis is open to competition at the "école doctorale Sciences de la Mer et du Littorale" (Marine and Coastal Sciences) scholarship, and success will initiate it, which will begin in September/October 2021 for 3 years.

Contact

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

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

References

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.