Institut méditerranéen de biodiversité et d'écologie - marine et continentale -

M2 Position in Cheminformatics

Artificial intelligence in metabolomics study by LC/MS and NMR of Natural Products

Plants produce and accumulate a wide variety of metabolites that can be of significant interest in various fields (particularly pharmaceutical and healthcare, agro-food, or cosmetics). The identification of these metabolites is the crucial first step in assessing their potential use. Traditional methods of compound identification involve purifying them, a lengthy and time-consuming process (taking up to several weeks to purify one molecule). Since an organism can accumulate a substantial number of molecules, the purification approach is not realistic for comprehensive characterization. Modern chromatographic and spectroscopic tools, such as liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) or nuclear magnetic resonance (NMR), nevertheless offer an alternative by accelerating the characterization of these molecules in complex matrices (extracts, fractions) without prior purification.

The main idea is to develop and train an AI model to systematically recognize the mass spectra of known natural molecules present in an extract, in order to identify them quickly. Additionally, the project aims to propose potential structures for unknown molecules. This will be achieved by annotating molecules from biological extracts *in silico* based on the spectra obtained by LC-MS and then validating these annotations through NMR analysis. The goal is to build a system for identifying molecules in complex biological extracts, combining the speed and cost-effectiveness of LC-MS analysis with the reliability of an AI trained model on real spectra, not just *in silico*-generated spectra, derived from chemical structures that have been rigorously validated by NMR.

The student will organize and expand a database of extracts and isolated natural origin molecules, with associated mass and NMR spectra acquired under standard conditions (concentration, MS, and NMR acquisition parameters) to train the AI tool to recognize spectra of molecules, whether they are isolated or in a mixture.

Requirements: Strongly motivated student with skills in cheminformatics and AI algorithms. The candidate should be interested in medicinal and organic chemistry, plants, and more globally to work at the interface between chemistry and informatics. The knowledge of Python language is required. She/he has to be curious, rapidly independent to conduct her/his own research, able to interact with several people and with teamwork spirit.

Application deadline : 20/10/2023

Internship period: 01/2024 – 07/2024

Internship location: Faculty of Pharmacy, 27 bd Jean Moulin, 1305 Marseille, France

Contacts:

Dr. Elnur Garayev – elnur.garayev@univ-amu.fr - IMBE

Pr. Sok-Siya Bun-Llopet - sok-siya.bun@univ-amu.fr - IMBE