## AI-Driven Exploration of Plant Chemical Space for Mass Spectral Elucidation

We invite applications for a post-doctoral research position focused on the development of algorithmic and AI-based methods for constrained graph generation, applied to chemical structure elucidation from experimental mass spectrometry data. This is an 18-month postdoctoral position, supervised jointly by experts in plant metabolite chemistry (LC-MS-based structural analysis) and AI methods for molecular enumeration.

The overall aim is to improve the identification of small molecules produced by plants, which occupy large, structurally diverse, and highly constrained graph spaces, where chemically valid structures constitute only a small fraction of all combinatorial possibilities. To address this, the project adopts a **building-block-based approach** in which molecules are modeled as **chemical graphs assembled from recurring substructures**, rather than atom-by-atom. These substructures correspond to chemically meaningful fragments frequently observed in plant metabolites, such as aromatic rings, glycosylated moieties, and organic acid units. From a computational perspective, this is formulated as a **constrained graph assembly and search task**, combining **symbolic AI, graph theory and knowledge-based reasoning**, and augmented with **machine learning** to extract fragment libraries and probabilistic priors from existing databases.

Key tasks and objectives:

- Analysis of plant metabolite databases using machine learning to detect recurring substructures, extract fragment libraries, and compute connection statistics or probabilistic priors.
- Represent these fragments as labeled graphs with explicit attachment points, and define symbolic bonding and compatibility rules based on chemical knowledge and graph-theoretic constraints to ensure chemical validity during assembly.
- Combinatorial generation of candidate molecular graphs using constraint programming (SAT and/or CSP), ensuring that structures (1) satisfy predefined graph and chemical constraints (valence, connectivity, isomorphism), (2) match molecular formula constraints inferred from high-resolution mass spectrometry, and (3) are compatible with experimental fragmentation patterns used as constraints or scoring criteria. Dedicated heuristics can also be designed and implemented for efficient exploration of the large combinatorial search space using isomorphism-free generation, canonical labeling, and pruning strategies to manage combinatorial explosion.
- Exploration and ranking of candidate structures within plant-relevant chemical space.
- Integration of developed algorithms into existing computational pipelines for mass-spectral structure elucidation.
- Benchmarking and comparative evaluation against reference datasets and existing structure-generation approaches.

Implementation will build upon and interface with established cheminformatics platforms that support fragment-based enumeration and rule- or reaction-based assembly, such as ChemLG [Afzal et al., 2019] and BuildAMol [Zhang et al., 2024], as well as RDKit with reaction SMARTS, while extending them with custom graph-theoretic algorithms and constraint-reasoning modules tailored to plant-derived chemical spaces.

**Location and Duration:**

The candidate will be attached to the MIS UR 4290 (https://mis.u-picardie.fr/) laboratory in Amiens and will work within the OCIA team. The position will last 18 months, starting in 2026, with a postdoctoral contract and a salary in line with French public-sector rates.

**Prerequisites:**

The candidate must be holding a PhD before the start of the postdoctoral project, with a strong background in AI, particularly combinatorial problem solving and constraint programming (Constraint Satisfaction and/or Satisfiability). Candidates do not need prior expertise in plant chemistry or mass spectrometry but any knowledge on cheminformatics or computational chemistry, especially molecular graph representations, would be appreciated.

Candidates should be motivated to develop new algorithms and tools for a real-world constrained graph generation problem, work in a multidisciplinary team, publish results in peer-reviewed journals, and participate in international conferences and workshops to present their work.

**Research Environment:**

The project is embedded in a collaborative framework linking the MIS (UR UPJV 4290) and BIOPI (UMRT BioEcoAgro INRAE 1158) laboratories of UPJV. The position is also supported by the Haut-de-France region through CPER CornelIA (https://project.inria.fr/cornelia/fr/).

The postdoctoral researcher will be jointly supervised by Dr Catherine Huyghe (MIS), expert in machine learning, Dr Sami Cherif (MIS), expert in symbolic AI and logic-based reasoning, and Dr Rebecca Dauwe (BioEcoAgro), expert in plant metabolite chemical structures, LC-MS-based chemical analysis, and computational metabolomics.

An office space, a computer, access to computing resources (MatricS UPJV platform) and funding to present in conferences will be provided.

**How to apply:**

To apply, please send the following documents by email:

- a curriculum vitae (CV)
- a motivation letter
- one recommendation letter (or contact details of a referee)

to sami.cherif@u-picardie.fr.

Review of applications will begin immediately and continue until the position is filled.

**Selected References:**


1. **Zhang, T., Sun, S., Wang, R., Li, W., Gan, B., & Zhang, Y.** (2024). *BuildAMol: a versatile Python toolkit for fragment-based molecular design.* Journal of Cheminformatics, 16**,** Article No. (2024). https://doi.org/10.1186/s13321-024-00900-6
2. **McKay, B. D., Yirik, M. A., & Steinbeck, C.** (2022). *Surge: a fast open-source chemical graph generator.* Journal of Cheminformatics, 14 (Article 24), 1-9. https://doi.org/10.1186/s13321-022-00604-9
3. **Carissan, Y., Hagebaum-Reignier, D., Prcovic, N., Terrioux, C., & Varet, A.** (2022): *How constraint programming can help chemists to generate Benzenoid structures and assess the local Aromaticity of Benzenoids*. Constraints An Int. J. 27(3): 192-248. https://doi.org/10.1007/s10601-022-09328-x
4. **Yirik, M. A., Sorokina, M., & Steinbeck, C.** (2021). *MAYGEN: an open-source chemical structure generator for constitutional isomers based on the orderly generation principle.* Journal of Cheminformatics, 13 (Article 48), 1-14. https://doi.org/10.1186/s13321-021-00529-9
5. **Afzal, M. A. F., Vishwakarma, G., Dudwadkar, J. A., & Hachmann, J.** (2019). *ChemLG - a library generator for the exploration and enumeration of chemical and materials spaces.* (Release 0.3). Hachmann Laboratory. https://github.com/hachmannlab/chemlg