Annual Report Year 1
Anticipating the Chemistry of Life (ACOL)

Marco Visani[1,2], Madleina Caduff[1,2], Daniel Wegmann[1,2], and Pierre-Marie Allard[1,2,3]

[1]Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland
[2]Swiss Institute of Bioinformatics, 1700 Fribourg, Switzerland
[3]Corresponding author, pierre-marie.allard@unifr.ch

November 4, 2025

Contents

1 Introduction

1.0.1 Context

1.0.2 Need

1.0.3 Task

1.0.4 Object

2 Markov Random Field

2.1 WP 1

2.1.1 Context

2.1.2 Need

2.1.3 Task

2.1.4 Object

The first months were dedicated for coaching me in C++. To do so, we started some simple implementations for the model. We started with some bitwise operations for the different storage vectors. Given the huge amount of data, we store in a single uint64 the state of the node (1 or 0), its coordinate, and the number of times it was a 1. This allowed me to understand some bitwise operations, how to create classes and the use of constructors.

$$\boldsymbol{\Lambda}_c = \begin{pmatrix} -\alpha_c & \alpha_c \\ 1 - \alpha_c & \alpha_c - 1 \end{pmatrix} \tag{1}$$

We changed from μ_0 and μ_1 to a single parameter α which is the change rate to go from 0 to 1.

$$\boldsymbol{P}(n) = \exp(\boldsymbol{\Lambda}_c \nu_c b(n)). \tag{2}$$

We now have an additional parameter ν_c which is the rate parameter of the clique.
Root state probabilities are given by the stationary distribution of the Markov chain:

$$\boldsymbol{P}_\infty = \left(1 - \alpha_c, \alpha_c\right). \tag{3}$$

$$\mathbb{P}(L_{ms}|\boldsymbol{x}(m,s), R_{ms}) = \begin{cases} \epsilon & \text{if } \boldsymbol{x}(m,s) = 0, L_{ms} = 1, \\ 1 - \epsilon & \text{if } \boldsymbol{x}(m,s) = 0, L_{ms} = 0, \\ R_{ms} & \text{if } \boldsymbol{x}(m,s) = 1, L_{ms} = 1, \\ 1 - R_{ms} & \text{if } \boldsymbol{x}(m,s) = 1, L_{ms} = 0. \end{cases} \tag{4}$$

Added an error rate ϵ in case a molecule was wrongly added to LOTUS.

$$R_{ms} = (1 - e^{-\gamma_0 P_m})(1 - e^{-\gamma_1 Q_s}) \tag{5}$$

Changed equation for research effort.

2.2 WP 2

2.2.1 Context

2.2.2 Need

2.2.3 Task

2.2.4 Object

Implementation of mass spec data still in discussion but we are planning to move from ViMMS [1] to just using outputs of annotation tools. We might use a combination of Sirius, Metfrag, CFM-ID and maybe weight these outputs.

2.3 Fake extract

2.3.1 Context

We would like to model the "loss" of molecules during the extraction process from Nature to the data analysis.

2.3.2 Need

2.3.3 Task

2.3.4 Object

Plate 3 was performed by a bachelor student, but results are weird, so we'll pass everything again. Plate 1 is done.

3 LOTUS Expanded

3.0.1 Context

Along with the PhD, there are some side projects. One of which is the in silico expansion of the molecules in LOTUS thanks to some reaction rules.

3.0.2 Need

This allows to TODO

3.0.3 Task

It turns out me generate some molecules that are true natural products so we would like to see if we can create this synergy between: in silco generation of a compound, create its in silico MS2 spectrum, try to detect them in real data, if we do, then expanded that molecule again.

3.0.4 Object

In silico chemical expansion of molecules in LOTUS, producing $3 \cdot 10^6$ molecules. Website is being build. We want to add the species to the visualization. Still need to figure out how this can be broadly used.

4 EMI-Monorepo

4.0.1 Context

EMI-Monorepo is our rust coded monolith were we are building all the tools for the Earth Metabolome Initiative. Our post doc Luca is leading the software development of the project. We are building everything in Rust from the backend to the frontend, to the data analysis part.

4.0.2 Need

Given the difficulty of the development of the portal, when Luca feels like a project could be a good training for me to learn Rust, he asks me to do it.

4.0.3 Task

4.0.4 Object

Helped Luca on the Monorepo for a couple of crates. Lost a bit the track of what he is doing.

5 Perspectives

5.0.1 Context

5.0.2 Need

5.0.3 Task

5.0.4 Object

I'd like to continue working with Luca in Rust as it is a skill that I want to master by the end of the PhD.

References

[1] J. Wandy et al. "ViMMS 2.0: A Framework to Develop, Test and Optimise fragmentation Strategies in LC-MS Metabolomics". In: *Journal of Open Source Software* 7.71 (30, 2022). DOI: 10.21105/joss.03990.