Internship project Master 2 Recherche
Year 2018 / 2019

Laboratory: UGA-IAB Inserm 1209, CNRS 5309 **Director:** Pierre Hainaut
Team: Equipe 8 **Head of team:** Patrice MARCHE
Name and status of scientist in charge of the project: Philippe BULET, DR CNRS, **HDR yes**

Host Laboratory Address:
Plateforme BioPark d'Archamps, Bâtiment le Forum, Archamps Technopole, 74160 Archamps
Phone: 0450432521**, e-mails:** philippe.bulet@univ-grenoble-alpes.fr, sebastien.voisin@biopark-archamps.org, dalel.askri@biopark-archamps.org

Title: Datamining of bee proteins identified by proteomics analysis in preparation to submission to an international protein sequence database

Objectives: To classify the peptides/proteins characterized in our laboratory during the proteomics LC-MS/MS analysis of hemolymph (insect blood) samples of bees and to assemble a list of peptide sequences with a high confidence of identification, with the aim to submit these sequences to a public database.

Abstract: Protein sequence databases used in proteomics are assembled by the accumulation of experimental data coming from different sources: genomic sequencing, transcriptomics, and direct experimental observation of the peptide sequences, such as through a high-resolution tandem mass spectrometry analysis (MS/MS). Protein annotation (name, function,…) for an unknown protein relies largely on searching sequence homologies with known proteins.
In the course of our laboratory's research projects, we acquired proteomics data on the hemolymph proteins of bumblebees and solitary bees. However, the available public protein sequences for these two non-model organisms are incomplete and not fully annotated. We have started to tentatively annotate the proteins we found by searching for sequence homologies inside NCBInr database with the BLAST tool.
We wish to further explore and valorize these obtained data, notably by assembling cases of peptide sequences backed by high-quality experimental data, and submitting these sequences for incorporation into a public protein sequence database. To achieve these aims, our obtained data will have to be mined for peptides characterized by MS spectra of the appropriate level of quality.

Methods: The tentative annotation of the non-annotated/uncharacterized proteins we identified in our samples will have to be completed, by using the NCBI BLAST homology search engine. The results of these searches will be used to update our laboratory's internal database of identified proteins. Next, the related peptide sequences of these non-annotated/uncharacterized proteins will have to be graded in function of the quality of the acquired MS/MS spectra. Finally, peptide groups passing a quality threshold will be proposed for addition to a public database.

Relevant publications of the team:
Masson V, **Arafah K, Voisin S, Bulet P**. Comparative Proteomics Studies of Insect Cuticle by Tandem Mass Spectrometry: Application of a Novel Proteomics Approach to the Pea Aphid Cuticular Proteins. Proteomics. 2018 Feb;18(3-4). doi: 10.1002/pmic.201700368
Pisani C, **Voisin S**, **Arafah K**, Durand P, Perrard MH, Guichaoua MR, **Bulet P**, Prat O. (2016) *Ex vivo* assessment of testicular toxicity induced by carbendazim and iprodione, alone or in a mixture. ALTEX. 2016;33(4):393-413.
Bulet P & Arafah K (2014) Proteomics in Methods in Immunology, from theory to practice Collège des enseignants d'immunologie (Assim) Coordinators: Marie-Christine Béné, Christian Drouet, Sylvain Fisson et Estelle Seillès

Requested domains of expertise (few keywords): Bioinformatics, database handling, proteomics