PhD Project Proposal (2023 – 2026)

Cellular functions of Micropeptides : a network approach

Short open reading frame (sORF)-encoded polypeptides (SEPs), also known as microproteins, have recently emerged as new key players in biology. Several microproteins were shown to be biologically active molecules and seem to be involved in a wide range of biological functions. They notably regulate many key cellular processes such as apoptosis, respiration, mTOR signaling, DNA damage repair and ER stress response. However, the cellular roles of thousands of SEPs are still unknown. As proteins seldom act alone, identifying their interacting partners informs on their cellular functions. The goal of this thesis work will therefore the discovery of the cellular functions of the SEPs identified in Drosophila through the inference and the analysis of their protein interaction network.

Background

Short open reading frame (sORF) encoded polypeptides, also called microproteins or SEPs (small encoded peptides), have recently emerged as new players in biology. SEPs, usually less than 100 amino acids in length, are produced from various RNAs: the 5' and 3' untranslated regions, introns of pre-mRNAs or alternative reading frames of the coding sequence of messenger RNAs. Long non-coding RNAs (lncRNAs) and rRNAs (ribosomal ribonucleic acids) were also found to code for SEPs. Several microproteins have been shown to be biologically active molecules and appear to be involved in a wide range of biological functions. In particular, they regulate many key cellular processes such as apoptosis, cellular respiration, mTOR signaling, DNA damage repair or the response to stress. However, the cellular roles of thousands of SEPs are still unknown.

Recently, our collaborators (S. Plaza & B. Fabre, LRSV, Toulouse) identified 401 SEPs in *Drosophila melanogaster* using deep peptidomics methods. By analyzing their sequences, we predicted the subcellular localization, evaluated the protein domains and short linear motifs (SLiMs) content for 235 of them, and highlighted the specific functions of these small proteins (Fabre et al., 2022).

Since proteins rarely act alone, the identification of their interaction partners provides information about their cellular functions. The objective of this thesis work will thus be the discovery of the cellular functions of the SEPs identified in Drosophila through the inference and the analysis of their protein interaction network.

Approach

The PhD student will infer the interactome between SEPs and canonical Drosophila proteins, using a computational framework called mimicINT that we have recently developed (Choteau et al., 2022), which relies on the detection of protein domains and SLiMs within SEPs (Zanzoni et al., 2017). The integration of these interactions into the overall Drosophila protein-protein interaction network will enable the investigation the functional modules of the network targeted by SEPs, an approach we have already successfully used in several previous analyses (Zanzoni et al., 2017; Kim et al., 2022), we hope to provide new insight into the functions of SEPs.
In addition, network analyses will be undertaken to better understand how certain regions of the network may be influenced by the functional and topological changes induced by SEPs. In order to identify these regions, we will use network propagation approaches, such the Random Walk with Restart algorithm. The resulting subnetworks will be further functionally investigated (e.g., by looking for over/under-representation of GO terms) to inform of SEP-induced network perturbations. Finally, some of the predictions will be experimentally validated by our collaborators.

Profile and skills

- Master in Bioinformatics or Systems Biology or Complex Systems or Data Science.
- Coding: Python, R.
- Experience in data analysis is a plus.

- As the project will be partially carried out in the context of an international collaboration, a good level of English is expected (B2, upper intermediate).

Contact information

Candidates can send inquiries to: Christine Brun (christine-g.brun@inserm.fr) and Andreas Zanzoni (andreas.zanzoni@univ-amu.fr).

Please, note that this is a competitive PhD call. The selected candidate will present the project in front of a jury of the doctoral school, which will rank applicants and assign a limited number of scholarships.

About the lab

The TAGC laboratory (Theories and Approaches to Genomic Complexity, http://tagc.univ-amu.fr), is a joint Inserm and Aix-Marseille University research unit located in the Luminy Campus, next to the well renowned “Les Calanques” National Park (Marseille, France). In the lab, we seek to decipher the biological mechanisms involved in the development and the physiology of an organism that, once disrupted, could lead to pathologies such as cancer, cardiomyopathies, infectious disease like malaria, or systemic inflammation. Over the years, the Network Biology team has developed algorithms and network-based approaches to understand the organization of cellular processes and protein multifunctionality by exploiting molecular interaction networks (protein-protein and protein-RNA).

References

- Fabre B, Choteau SA, Duboé C, Pichereaux C, Montigny A, Korona D, Deery M.J, Camus M, Brun C, Burlet-Schiltz O, Russell S, Combier JP, Lilley K.S. and Plaza S. (2022) In depth exploration of the alternative proteome of Drosophila melanogaster. Front Cell Dev Biol, 10:901351.
- Choteau SA, Cristianini M, Maldonado K, Drets L, Boujeant M, Brun C, Spinelli L, Zanzoni A. (2022) bioRxiv 2022.11.04.515250; doi: https://doi.org/10.1101/2022.11.04.515250.
- Zanzoni A, Spinelli L, Braham S, and Brun C. (2017) Perturbed human sub-networks by Fusobacterium nucleatum candidate virulence factors. Microbiome, 5: 89.
- Kim DK, Weller B, Lin CW, Sheykhkarimli D, Knapp JJ, Dugied G, Zanzoni A. et al. (2023) A proteome-scale map of the SARS-CoV-2 human contactome. Nature Biotech. doi: 10.1038/s41587-022-01475-z.