Title: *Can metabolism process information?*

Supervisor: Jean-Loup FAULON, Professor and Director of Research

Institution: MICALIS, INRA

Address: Bat. 442, Domaine de Vilvert, 78352 Jouy en Josas Cedex, France

Description of the team (people with their position, main topics, website):

The team, Bio-RetroSynth is working the fields of systems and synthetic biology. Within these fields we have specialized in understanding and engineering metabolism. We are known for our Retro-synthesis and Machine Learning developments. The team comprises 13 staff members divided between a dry and a wet lab. Dry lab: 1 Professor, 1 Research Engineer, 1 Invited Senior Scientist, 3 Post-docs, 1 PhD student. Wet lab: 1 Researcher, 1 Associate Professor, 1 Post-doc, 3 PhD students.

www.jfaulon.com

Tel: 06 17 47 51 51, email: Jean-Loup.Faulon@inra.fr

Project Summary:

Can we engineer a perceptron (basic block of neural network architecture) in a microorganism? Conversely, can we use cellular network motifs as architectures for deep neural network?

Prior answering these questions, we first need to probe if and which information processing modules exist in organisms and in particular in microorganisms we can engineer. This is the main goal on the internship.

More precisely, the recruited student will search information processing motifs (logic gates, bi-fans, monolayer and multi-layer perceptrons, recurrent loop) in metabolic databases to probe to which extent metabolism plays a role in transducing, integrating and processing signals. Working with metabolism is relevant to microorganism in which signaling networks are not as sophisticated as in higher organisms.

During his internship, the student will benefit from the team expertise in metabolism and information processing devices as well as from the many tool (RetroPath [1, 2], Sensipath [3][4], RetroRules (retrorules.org) we have developed the past few years.

Description:

Quorum molecules, hormones, neurotransmitters and extracellular chemicals produce signals that are transduced down to the genetic layer in living organisms. Information is generally passed on through cellular receptors then signaling pathways to ultimately produce a transcriptional response.

In most instances, transcriptional activation requires the multiplexed detection and integration of different signals. For instance, bacteria like *Pseudomonas aeruginosa* are capable of detecting multiple quorum molecules to regulate virulence and biofilm formation through multi-layered

signaling networks [5]. More sophisticated are the complex decision-making molecular circuits (similar to perceptron) found in plants to control responses to a large range of environmental stimuli [6], interconnected together to what has been named the wood wide web [7].

Enzymes are key players in signal transduction as they are the catalyst of the pathways. Since enzyme produce metabolites, there are also instances in which signals are propagated through metabolism [8]. The best-known example is within the lac operon, where the transcription factor *LacI* is activated by allolactose, a metabolite produced from lactose by a secondary activity of beta-galactosidase [9]. The underground activity of β-Galactosidase is present only when *LacI* is present and this network motif is conserved by evolution [9].

The goal of the internship will be to probe to which extent metabolism plays a role in transducing, integrating and processing signals in particular with bacteria where signaling pathways are not as developed as in higher organisms.

The work will consist of searching motifs in metabolic databases corresponding to signal transduction, integration and processing. For signal transduction one will have to search for pathways linking metabolites to transcription factors or riboswitches, for signal integration and processing one might consider logic gates as it has been suggested in the past [10] or analog electronic motifs such as (weighted) adder, substractor, attenuator, perceptron, and recurrent loop, some which already found in signaling pathways [11]. **To validate the motifs found in metabolic maps of specific organisms, the work should yield experimental designs** to be carried by the team members of the recruiting group and/or during a follow-on PhD.

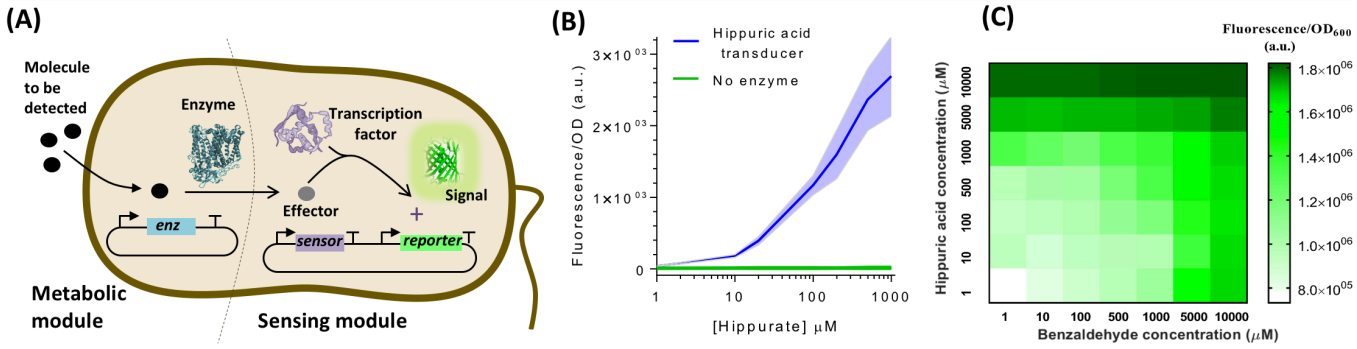

Figure 1. Example of experimental work to be carried out to validate a weighted adder. ***(A)*** *Metabolic and sensing modules. The metabolic module transforms (here via an hydrolase see* [12]) *a non-detectable molecule (hippurate) into an effector (benzoate). The sensing module comprises a transcription factor (BenR), which once bounded to the effector (benzoate) actives the expression of a report gene (RFP).* ***(B)*** *The dose−response curves (RFP fluorescence fold change vs. concentration of the molecule to be detected) are shown for engineered E. coli strains with and without the transforming enzyme (hydrolase).* ***(C)*** *An adder where hippurate and benzaldehyde are both transformed into benzoate through hydrolases, benzoate in turns activates RFP expression via BenR. Fluorescence is function of the weighted sum of the concentrations of hippurate and benzaldehyde.*

The recruited student will benefit from preliminary work we have carried out exhibiting

examples beyond allolactase and *LacI* where metabolic reactions are transducing signals [8, 13] along with a database [4] and a software tool (SensiPath [3]) we have compiled for sensing enabling metabolic pathways. The recruited student will work closely with two PhD students (Mathilde Koch and Amir Pandi) and an IT engineer (Thomas Duigou) who are currently designing and engineering signal transduction, integration and processing using metabolic pathways in the context of whole cell and cell-free biosensors.

Preliminary Work plan outline for a 6-month internship

- **Task 1**. Literature survey and motif search code benchmarking (1 month). Review literature on motif detection (applicable to metabolic networks). Benchmark a code written by a former master student (Ivan Valiev, Skoltech Institute, Moscov) to search for motifs in metabolic networks (using Kegg maps for instance).
- **Task 2.** SensiPath Database update (1 months). You will learn the tools we have developed to build the database (mostly the RetroPath2.0 [1, 2] workflows available at MyExperiment.org), you'll then generate an updated database with these tools.
- **Task 3**. Motif search in updated SensiPath database (3 months). Here you will have first to work out the topology of the motifs to be searched. Potential motifs are transducer, actuator, adders, 1 layer or multilayer perceptron, recurrent loops, bi-fan, and logic gates. Once the motifs will have been designed you will search them on the database build in Task 2 and with the tools developed in Task 1. Most of the work here will be to determine the topology of the motifs, as the search itself should be straightforward. You will also propose plan of experiments for experimental validation of the motifs found on SensiPath Database. There will be a possibility to carry out experimental work during a follow-on PhD.
- **Task 4.** Complete the work started in Task 3 and write report suitable for publication (1 month).

Follow-on work through a PhD program

There will be a possibility to continue the work carried out during the internship though a PhD scholarship. The PhD subject should be around learning network architecture *in silico* and *in vivo*. For instance, can we learn deep neural network architecture from biology? and what engineering biology can borrow from machine learning architecture and algorithms?

References (from the team *)

1. *Delepine, B.*, et al.*, RetroPath2.0: A retrosynthesis workflow for metabolic engineers. ***Metab Eng,*** 2018. 45: 158-170.
2. *Koch, M.*, et al.*, Molecular structures enumeration and virtual screening in the chemical space with RetroPath2.0. ***J Cheminform,*** 2017. 9(1): 64.
3. *Delepine, B.*, et al.*, SensiPath: computer-aided design of sensing-enabling metabolic pathways. ***Nucleic Acids Res,*** 2016. 44(W1): W226-31.
4. *Koch, M.*, et al.*, A dataset of small molecules triggering transcriptional and translational cellular responses. ***Data Brief,*** 2018. 17: 1374-1378.

5. Lee, J.*, et al.*, The hierarchy quorum sensing network in Pseudomonas aeruginosa. ***Protein Cell,*** 2015. 6(1): 26-41.

6. Scheres, B.*, et al.*, The plant perceptron connects environment to development. ***Nature,*** 2017. 543(7645): 337-345.

7. Helgason, T.*, et al.*, Ploughing up the wood-wide web? ***Nature,*** 1998. 394(6692): 431.

8. *Libis, V.*, et al.*, Sensing new chemicals with bacterial transcription factors. ***Curr Opin Microbiol,*** 2016. 33: 105-112.

9. Wheatley, R.W.*, et al.*, Structural explanation for allolactose (lac operon inducer) synthesis by lacZ beta-galactosidase and the evolutionary relationship between allolactose synthesis and the lac repressor. ***J Biol Chem,*** 2013. 288(18): 12993-3005.

10. Arkin, A.*, et al.*, Computational functions in biochemical reaction networks. ***Biophys J,*** 1994. 67(2): 560-78.

11. Alon, U., *An Introduction to Systems Biology: Design Principles of Biological Circuits*. 2006: Chapman and Hall/CRC.

12. *Libis, V.*, et al.*, Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways. ***ACS Synth Biol,*** 2016.

13. *Libis, V.*, et al.*, Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways. ***ACS Synth Biol,*** 2016. 5(10): 1076–1085.

Additional questions:

Does this project constitute the first steps of a PhD thesis that will be supported by a PhD fellowship? **YES**

Do you have any special accommodation or fellowship for foreign students? **YES. Apartments for foreign students are available near INRA Jouy-en-Josas campus.**