UNIVERSITÉ DE RENNES 1

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2021

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : MATHSTIC **POLE DE RECHERCHE : MATHSTIC**

Unité de recherche d'accueil

Unité de Recherche : IRISA

Equipe : Dyliss

Encadrement

Directeur de thèse : SIEGEL **Unité de recherche** : Anne

Employeur/Établissement d'origine : CNRS

HDR : ■ oui □ non **Nb de thèses en cours** : **1**

Mail : anne.siegel@irisa.fr **Tél**. : 0299847448

Co-directeur de thèse : Paulevé Loic **Unité de recherche** : LABRI (Bordeaux)

Employeur/Établissement d'origine : CNRS

HDR : ■ oui □ non **Nb de thèses en cours** : 1

Mail : loic.pauleve@labri.fr **Tél.** :

Sujet

Titre : Scalable inference of Boolean rules controlling hybrid models of multi-layer biological systems.

Mots clés : bioinformatics, systems biology, abstract interpretation, declarative programming, hybrid modeling.

Cotutelle internationale (non) : non

Sollicitation d'un cofinancement EDUC (oui/non) : non

Descriptif :

This thesis aims at developing new analytical formalisms for inferring molecular regulatory networks that control metabolism in order to elucidate from "omics" data how regulations mediate metabolic mechanisms (and conversely).
Biology has motivated computer scientists and mathematicians to develop specific formalisms to capture the characteristics of intracellular behaviors. Cellular regulation is known to be multi-layer. Regulatory and signaling interactions are highly nonlinear, modeled with non-deterministic frameworks, from logical to rule-based models [AT14,PKCH20]. These models are identified from transcriptomic or phosphoproteomics data by solving combinatorial or MILP problems in order to optimize data-fitting and parsimony hypotheses. Designing a scalable solution to this identification problem was done in [VGE15,OPS16] by relying on abstract interpretation and declarative programming.
The metabolic layer faces a scalability issue since models involves several thousands of compounds and interactions. This prevents using ODE models whose kinetic parameters cannot be identified uniquely. Modelers usually assume that internal metabolites are in a quasi-steady-state, as in the dynamic enzyme-cost Flux Balance Analysis (deFBA) [WOB15] which models metabolism by a dynamic optimization problem, without knowing the underlying regulatory mechanisms.
The combined metabolic and regulatory levels are necessary to figure out how gene expression triggers specific phenotypes depending on the environmental constraints. Most of the approaches integrating metabolic and regulatory networks are based on Boolean logic, such as *regulatory FBA* (rFBA) [CSP01,M15] or *steady-state regulatory FBA* (SR-FBA) [SE07]. These approaches include Boolean rules in the optimization process of FBA using mixed-integer linear programming (MILP), and finds a steady-state for both the metabolic and the regulatory network. In this case, the metabolic layer is inferred from genomic and experimental data by solving combinatorial or MILP optimization problems based on quasi-steady state hypotheses**.** However, the transcriptional and signaling layers are manually curated from the literature or experimental data evidence.
The goal of the phD will be to extend the methods developed for the identification of Boolean models developed in [VGE15,OPS16] in order to enable a scalable identification of multi-layer models and elucidate from "omics" data how regulations mediate metabolic mechanisms (and conversely). The main drawback of in [VGE15,OPS16] is that they all rely on a simplified assumption on the system functioning, allowing to focus only on a single (regulatory/signalling) layer of the biological system, adjoined with several dynamical assumptions (steady-state, synchronous dynamics, possibly time-series and asynchronous when using abstract interpretation techniques) [PKCH20]. In this case, interactions between layers are totally neglected.

The main issues to address during the phD are (1) contribute to the design of a Hybrid framework for metabolic regulation. This will be done by extend the deFBA to a hybrid discrete-continuous framework *regulatory deFBA* that will include regulations as discrete jumps between regulatory states, each having a specific continuous dynamics given by differential-algebraic equations [LB19]. This approach underlines models in [RKBS17] but requires a formalization. (2) Designing a new the inference problem on the hybrid framework, by extending the problems solved in [VGE15,OPS16] (3) Solve the optimization problems with combinatorial, linear and temporal features. In this case, we will rely on recently introduce temporal extensions of Answer Set Programming (Tellingo), possibly by designing specific temporal extension of ASP with constructs from linear temporal logic and (linear) dynamic logic [JKO17,BCD18] as done for solving previous optimization problems in systems biology [FSS19].

[AT14] Albert R, Thakar J. *Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions*. Wiley Interdiscip Rev Syst Biol Med. 2014

[BCD18] A. Bosser, P. Cabalar, M. Dieguez, and T. Schaub. *Introducing temporal stable models for linear dynamic logic*. In KR'18. AAAI Press, 2018

[CSP01] Covert MW, Schilling CH, Palsson B. *Regulation of gene expression in flux balance models of metabolism*. J Theor Biol. 2001

[FSS19] Frioux C, Schaub T, Schellhorn S, Siegel A, and Wanko P. *Hybrid metabolic network completion*. Theory and Practice of Logic Programming. 2019.

[JKO17] anhunen T, Kaminski R, Ostrowski M, Schaub, Schellhorn S, and Wanko P. *Clingo goes linear constraints over reals and integers* TPLP. 2017.

[LB19] Liu, L. and Bockmayr, A. Formalizing metabolic-regulatory networks by hybrid automata. bioRxiv (2019).

[M15] Marmiesse, L., Peyraud, R., and Cottret, L. (2015). Flexflux: Combining metabolic flux and regulatory network analyses. BMC systems biology, 9:93.

[OPS16] Ostrowski M, Paulevé. L, Schaub T, Siegel A, Guziolowski C. Boolean network identification from perturbation time series data combining dynamics abstraction and logic programming. BioSystems. 2016

[PKCH20] Pauleve, L., Kolcak, J., Chatain, T., and Haar, S. (2020). Reconciling Qualitative, Abstract, and Scalable Modeling of Biological Networks. Nature Communications, 11.

[RKBS17] Reimers AM, Knoop H, Bockmayr A, Steuer R. *Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth*. PNAS 2017

[SE07] Shlomi T, et al *A genome-scale computational study of the interplay between transcriptional regulation and metabolism.* Mol Syst Biol. 2007

[VG515] Videla S et al, Learning Boolean logic models of signaling networks with ASP, Theoretical Computer Science, 2015.

[WOB15] Wladherr S, Oyarzun DA, Bockmayr A, *Dynamics optimization of metabolic networks coupled with gene expression*. J. Theor. Biol. 2015.