

# Calculer des plans expérimentaux pour contrôler des systèmes biologiques à l'aide de programmation logique

3 février 2015

**Receiving team :** MeForBio

**Research laboratory :** IRCCyN, Ecole Centrale de Nantes

**Website :** <http://www.irccyn.ec-nantes.fr/meforbio>

**Student mentor :** Carito Guziolowski ([carito.guziolowski@irccyn.ec-nantes.fr](mailto:carito.guziolowski@irccyn.ec-nantes.fr))

## Description du projet

Les réseaux de signalisation sont des mécanismes intracellulaires qui peuvent être modélisés par différentes approches logiques : réseaux booléens, réseaux logiques asynchrones, process hitting (algebraic modelling). Dans tous les cas, des données expérimentales ou issues de la littérature fournissent des informations sur l'état de certaines protéines après que différentes perturbations des entrées du système se sont propagées via des interactions jusqu'à une stabilisation du système. On dispose maintenant de méthodes pour apprendre comment les molécules doivent s'associer (sous la forme de règles logiques) pour transcrire correctement les réponses observées dans des données réelles. La plupart du temps, il s'avère cependant qu'il y a en fait beaucoup de modèles ou réseaux qui sont capables d'expliquer tout aussi bien les données et les connaissances sur des systèmes.

Lorsque ces familles de modèles sont apprises, deux types de travaux sont proposés, en terme d'expérimentations :

- Certains cherchent à proposer des expérimentations pour forcer le système à réaliser des comportements attendus ou prédits (contrôle) [7, 6, 2].
- D'autres cherchent à proposer des expérimentations qui permettent de réduire le nombre de modèles compatibles et mieux cerner le comportement des voies de signalisation, et ensuite proposer des expérimentations pour forcer le système à réaliser des comportements attendus (design expérimental) [1].
- D'autres enfin cherchent à améliorer les modèles en essayant d'y rajouter des interactions qui avaient été mal identifiées a priori [3].

L'équipe MeForBio, en association avec l'équipe Dyliss (IRISA - Rennes), a proposé des modélisations sous la forme de problèmes d'optimisation combinatoire pour résoudre les questions du design expérimental et du contrôle du système [6]. Ces problèmes d'optimisation sont résolus à l'aide de programmation logique en utilisant des approches de Programmation par Ensembles Réponses (approche déclarative associée à des solveurs de type SAT) [4]. Les méthodes ont été intégrées dans un package python nommé caspo (<http://bioasp.github.io/caspo/>) [5].

Cependant, les applications sur des données réelles montrent que les approches de design expérimental ne sont pas assez performantes sur certains jeux de données réels, parce que les espaces de recherche deviennent trop gros à explorer et les informations sur le système trop grossières.

L'objectif de ce travail de recherche sera de comprendre quels sont les enjeux du *experimental design*, et de faire un état de l'art sur comment ce problème est approché par des méthodes informatiques actuels. Noter que le *experimental design* peut aussi se confondre avec la notion de recherche des paramètres dans des modèles à base des équations différentielles. En illustrant pour quoi une approche par programmation logique s'avère opportune pour ce type de problème, quelles seront ses faiblesses et ses qualités par rapport aux approches existantes. En termes d'implémentation il faudra partir des briques existantes (caspo-design, caspo-control, caspo-learning) pour proposer une nouvelle modélisation du design expérimental en introduisant des contraintes supplémentaires pour réduire les espaces de recherche à explorer. En particulier, il sera intéressant de rechercher les expérimentations qui permettent à la fois de réduire l'ensemble des modèles compatibles et de forcer le système à avoir certains comportements attendus. Il faudra de plus proposer les meilleurs critères pour valider ces approches sur des données simulées puis réelles.

## Références

- [1] Atias, N., Gershenson, M., Labazin, K., and Sharan, R. (2014). Experimental design schemes for learning Boolean network models. *Bioinformatics*, **30**(17), i445–i452.
- [2] Dunn, S. J., Martello, G., Yordanov, B., Emmott, S., and Smith, A. G. (2014). Defining an essential transcription factor program for naive pluripotency. *Science*, **344**(6188), 1156–1160.
- [3] Eduati, F., De Las Rivas, J., Di Camillo, B., Toffolo, G., and Saez-Rodriguez, J. (2012). Integrating literature-constrained and data-driven inference of signalling networks. *Bioinformatics*, **28**(18), 2311–2317.
- [4] Gebser, M., Kaminski, R., Kaufmann, B., and Schaub, T. (2012). Answer Set Solving in Practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **6**(3), 1–238.
- [5] Guziolowski, C., Videla, S., Eduati, F., Thiele, S., Cokelaer, T., Siegel, A., and Saez-Rodriguez, J. (2013). Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming. *Bioinformatics*, **29**(18), 2320–2326.
- [6] Kaminski, R., Schaub, T., Siegel, A., and Videla, S. (2013). Minimal intervention strategies in logical signaling networks with asp. *TPLP*, **13**(4-5), 675–690.
- [7] Sharan, R. and Karp, R. M. (2013). Reconstructing Boolean Models of Signaling. *Journal of Computational Biology*, **20**(3), 249–257.