

# Enrichissement des méthodes d'analyse dynamique des Réseaux de Régulations Biologiques à l'aide de la programmation logique

## Contexte

L'étude des mécanismes du vivant à différentes échelles est aujourd'hui rendue possible par les progrès de l'outil informatique. La compréhension de la machinerie cellulaire, et notamment des interactions entre gènes et protéines, permet le développement de thérapies géniques.

À ces fins, il est nécessaire d'établir des modèles à la fois robustes et efficaces, qui permettent de s'affranchir du manque potentiel de données expérimentales. Différents modèles ont été proposés afin de représenter les influences entre composants de la machine cellulaire tout en respectant ces critères. Le Process Hitting (PH) [1] permet une représentation très atomique de ces systèmes d'interactions et a été développé au sein de l'équipe MeForBio. Ce formalisme offre des possibilités d'analyse statique puissantes permettant d'obtenir des résultats concrets sur des questions d'atteignabilité au sein de modèles de grande taille, là où des modélisations plus classiques se heurtent à des problèmes de temps de calcul ou de taille en mémoire.

Pint [2] est une bibliothèque écrite en OCaml permettant de lire des modèles au format Process Hitting, et de réaliser sur ceux-ci certaines opérations d'analyse développées formellement. Il s'avère cependant que pour des opérations de complexité importante, d'autres outils performants existent et permettraient d'obtenir des résultats de façon plus efficace.

L'*Answer Set Programming* (ASP) [3], est un langage de programmation logique [4] qui s'inspire de techniques SAT, s'est notamment révélé utile afin de gérer des problèmes de complexité exponentielle. Lors de précédents travaux au cours d'un précédent travail de master [6], des méthodes ont été développées pour l'analyse des propriétés dynamiques telle la recherche des états stables. De plus la méthode développée pour répondre à la question d'atteignabilité permet d'obtenir des résultats concrets au sein de modèles représentés en Process Hitting. Ces modèles représentent des réseaux de régulation biologiques de grande taille là où des modélisations plus classiques que PH se heurtent à des problèmes de temps de calcul ou de taille en mémoire. Le sujet de master porte sur l'enrichissement de ces méthodes programmées en ASP afin d'obtenir à terme des performances supérieures aux méthodes basées sur des paradigmes plus traditionnels pour des modèles de très grande taille. Nous désirons notamment étendre les méthodes à un modèle plus expressif de Process Hitting, permettant de représenter des priorités ainsi que des délais entre les actions (interactions entre les composants) modélisant la dynamique du système [8], mais aussi envisager un spectre plus large de propriétés que les seuls points fixes et problématiques d'atteignabilité.

## Séminaire bibliographique:

Ce travail nécessite une étude bibliographique sur le thème suivant :

- **Titre :**

« La programmation logique pour la bioinformatique »

## - Encadrants :

Emna Ben Abdallah ([emna.ben-abdallah@irccyn.ec-nantes.fr](mailto:emna.ben-abdallah@irccyn.ec-nantes.fr) )

Morgan Magnin ([morgan.magnin@irccyn.ec-nantes.fr](mailto:morgan.magnin@irccyn.ec-nantes.fr) )

## - Contexte :

Dans la perspective de l'analyse de systèmes modélisant des phénomènes de régulation biologique, la programmation logique s'est révélée bien adaptée pour faire face à la complexité des modèles étudiés à grandes échelles[7][8]. Le travail de ce séminaire bibliographique consistera à découvrir les concepts essentiels de la programmation logique (notamment l'*Answer Set Programming (ASP)* [3][4][6], implémentée à travers différents solveurs dont clingo [5]) et son application dans le domaine de la bioinformatique.

## Travail attendu

- Recherche bibliographique sur les méthodes de modélisation formelle de réseaux de régulation biologique, et notamment sur le Process Hitting,
- Prise de connaissance des outils et langages de programmation logique, et notamment d'ASP,
- Étude des méthodes d'analyse dynamique relatives au Process Hitting en ASP et leur enrichissement par la simplification de la modélisation des graphes : remplacer les sortes coopératives par des actions plurielles [6] et l'ajout des priorités et des délais dans les actions.
- Choix des outils adaptés, implémentation et analyse des performances.

## Environnement et cadre de travail

L'équipe MeForBio (Méthodes Formelles pour la Bioinformatique) de l'IRCCyN (Institut de Recherche en Communications et Cybernétique de Nantes) est composée de trois permanents, cinq doctorants. Elle est impliquée dans plusieurs projets d'envergure nationale dont HyClock (ANR) et GRIOTE (projet fédérateur régional) et BioTempo (ANR), et possède des partenariats nationales et internationales (notamment avec l'Allemagne et le Japon).

Le sujet de recherche s'intègre dans une optique d'enrichissement et d'amélioration de la bibliothèque Pint qui est en adéquation avec les objectifs de l'équipe MeForBio concernant l'étude de modèles de grande taille.

## Prérequis

Aucun prérequis en biologie n'est nécessaire pour effectuer ce stage de Master. Des connaissances préliminaires en programmation logique (ASP) ou en programmation par contraintes seraient un plus, mais ne sont pas requises.

## Références

[1] Loïc Paulevé, Morgan Magnin and Olivier Roux. Refining dynamics of gene regulatory networks in a stochastic  $\pi$ -calculus framework. *Transactions on Computational Systems Biology XIII*, volume 6575 of *Lecture Notes in Computer Science*, pages 171-191. Springer Berlin / Heidelberg, 2011. <http://loicpauleve.name/PMR10-TCSB.pdf>

[2] <http://loicpauleve.name/pint/>

[3] Chitta Baral. Knowledge Representation, Reasoning and Declarative Problem Solving. Cambridge University Press, 2003.

[4] Chitta Baral and Michael Gelfond. Logic programming and knowledge representation. *Journal of Logic Programming*. 1994, Vol. 19, 73-148.

[5] <http://potassco.sourceforge.net/>

[6] Emna Ben Abdallah. Recherche exhaustive des propriétés dynamiques dans le Process Hitting en utilisant l'Answer Set Programming. Thèse de Master, École Centrale de Nantes, septembre 2014. [http://www.irccyn.ec-nantes.fr/~benabdal/These\\_Master.pdf](http://www.irccyn.ec-nantes.fr/~benabdal/These_Master.pdf)

[7] Maxime Folschette, Loïc Paulevé, Katsumi Inoue, Morgan Magnin, and Olivier Roux. Identification of Biological Regulatory Networks from Process Hitting models. *Theoretical Computer Science*, December 2014. In press. Extended version of the article published in CMSB'12. [http://maxime.folschette.name/Folschette\\_TCS14.pdf](http://maxime.folschette.name/Folschette_TCS14.pdf)

[8] Maxime Folschette. Modélisation algébrique de la dynamique multi-échelles des réseaux de régulation biologique. Thèse de Doctorat, Ecole Centrale de Nantes, octobre 2014. [http://maxime.folschette.name/manuscrit\\_Folschette.pdf](http://maxime.folschette.name/manuscrit_Folschette.pdf)