

Sujet de thèse en informatique (bioinformatique)

Contexte

La biologie synthétique est un axe de recherche en plein essor dans le domaine de la conception et de l'ingénierie de systèmes basés sur des règles fonctionnelles biologiques dans le but d'obtenir de nouvelles fonctionnalités qui ne sont pas présentes dans la nature. D'après le MIT, la biologie synthétique est focalisée sur la conception raisonnée de systèmes biologiques, plutôt que sur la compréhension de la biologie des organismes naturels, c'est en ce sens qu'elle se distingue de la biologie classique.

On peut distinguer trois types d'approches de la biologie synthétique :

- *in vivo* où on utilise l'ingénierie génétique, par exemple, pour optimiser dans un organisme vivant, une voie métabolique produisant un composé chimique d'intérêt thérapeutique ou bien industriel. [1]
- *in vitro* dans laquelle on conçoit des composants artificiels pour des applications de nano technologies où ces composants sont conçus et déployés en dehors de cellules vivantes.
- *de novo*, qui est le genre le plus extrême de nano technologie bio-inspirée où des systèmes biologiques complètement artificiels proviennent d'une théorie *top-down* pour obtenir à partir d'assemblages de composants artificiels ce qui peut être considéré comme un système vivant (*Vie Artificielle*). [2]

Le projet BS² du laboratoire SysDiag à Montpellier, a un volet *Biologie Synthétique* dans lequel collabore P. Amar de l'équipe BIOINFO du LRI. Cette partie du projet utilise la technique de la biologie synthétique *de novo* dans le but d'obtenir un calculateur biologique embarqué (*bio-computeur*) dédié au diagnostic médical de pathologies ciblées.

Ce nanosystème biologique prendra la forme d'une vésicule (liposome) composée d'un réseau enzymatique biologique (métabolique) conçu artificiellement *ab initio* afin de servir à la fois de senseur des différents biomarqueurs mais aussi de calculateur (algorithme logique) et de révélateur (production d'une coloration par exemple). Les vésicules et le réseau biologique artificiel seront principalement composés d'enzymes et de métabolites d'utilisation courante en biotechnologie, donc stables et robustes.

La conception de ce nanosystème biologique artificiel nécessite 4 grandes étapes :

1. Conception *in silico* du réseau biologique artificiel
2. Modélisation et simulation dynamique du système pour l'optimisation et les études de robustesse,
3. Construction expérimentale et validation analytique du fonctionnement du système
4. validation *in-vivo*

Ces différentes étapes sont intimement liées et nécessitent bien entendu le travail conjoint de biochimistes, modélisateurs (informaticiens) et de cliniciens.

La thèse proposée ici s'inscrit dans le cadre des deux premières étapes du processus.

Conception du système

La conception du réseau artificiel se fera dans un premier temps sur la base d'un réseau biologique fonctionnel, abstrait à partir d'actions élémentaires [3] combinées entre elles et reliées par les substrats et les produits. Ce réseau abstrait sera construit et implémenté avec des enzymes spécifiques en utilisant le logiciel de design de réseaux biologiques développé à SysDiag : BioNetCad [4]. Cet outil permet notamment de réaliser l'implémentation du réseau biologique artificiel abstrait par satisfaction de contraintes. Les biomarqueurs d'intérêt, qui sont en fait des enzymes, feront partie de

ce réseau artificiel de manière à ce que leur présence/absence ou variation en quantité aient une influence sur le fonctionnement général du système. Cette influence sera notamment implémentée par une sous-structure du réseau jouant le rôle de circuit logique chargé de calculer une réponse intégrée (algorithme de décision) du système diagnostic.

Modélisation, simulation et validation

Cette étape de conception est suivie d'une étape d'évaluation fonctionnelle *in silico* s'appuyant sur des approches de modélisations des systèmes biologiques et de simulation dynamiques. Pour tester qualitativement et quantitativement notre système en amont des tests *in vitro*, nous utiliserons, outre les méthodes de type ODE, le logiciel HSIM [5], développé au LRI, pour effectuer des simulations numériques. Ce logiciel est un automate stochastique qui utilise une technique dite "entité-centrée" pour simuler le comportement et les interactions de bio-molécules dans un compartiment délimité par une membrane. Le simulateur étant discret et stochastique, il permet de calculer précisément le comportement de modèles ayant très peu de copies de certaines molécules. HSIM rend aussi compte de la localisation spatiale des molécules dans le compartiment et peut donc intégrer des phénomènes d'inhomogénéité dans la distribution spatiale des molécules dans la vésicule et de leurs conséquences sur la dynamique du système.

Objectifs de la thèse

Le but de cette thèse est de systématiser à l'aide d'outils informatiques et biotechnologiques la conception, l'implémentation, le test et la validation de bio-computeurs dans le contexte du diagnostic médical.

Les objectifs sont multiples :

- l'étude d'outils formels d'extraction automatique de circuits logiques fonctionnels à partir des réseaux métaboliques qui se trouvent dans des organismes vivants pour en constituer une bibliothèque réutilisable
- l'étude de méthodes intégrées pour la conception de bio-calculateurs, leur implémentation compte tenu de diverses contraintes (robustesse, stabilité, non toxicité, etc.) et de leur validation *in silico* en utilisant les approches les plus adéquates (résolution de systèmes d'équations différentielles, simulations stochastiques "entité-centrée" ou globales)
- enfin, la thèse aura un volet *application* consistant à l'issue des réflexions précédemment menées à réimplémenter une nouvelle version d'un outil de type BioNetCad, à la fois plus dédiée à notre approche et plus intégrée aux outils de simulations.

Encadrement

La thèse s'effectuera au LRI, avec des séjours au laboratoire SysDiag à Montpellier. Elle sera co-encadrée par Patrick AMAR, de l'équipe *BIOINFO* du LRI / équipe-projet *AMIB* de l'INRIA et par Franck MOLINA, directeur du laboratoire SysDiag (UMR CNRS 3145) à Montpellier.

Références

- [1] John McCaskill, Kristian Lindgren *et al.* Programmable Artificial Cell Evolution (PACE) In *Modelling Complex Biological Systems in the Context of Genomics*, pages 27–29. EDP Sciences - isbn : 978-2-7598-0019-3, 2007.
PACE project website : <http://www.istpace.org>
- [2] Pier Luigi Luisi, Pascale Stano *et al.* En route to semi-synthetic minimal cells. In *Modelling Complex Biological Systems in the context of genomics*, pages 19–30. EDP Sciences - isbn : 978-2-7598-0014-8, 2006.
SynthCells project website : <http://www.syntcells.org>

- [3] Sabine Peres, Stephanie Rialle, Liza Felicori and Franck Molina. Formal language for detailed structure function annotation based on elementary bricks of action *Bioinformatics*, 26(12) :1542–1547, 2010.
- [4] Stephanie Rialle, Liza Felicori, Camila Dias-Lopes, Sabine Peres, Sanaa El Atia, Alain R. Thierry, Patrick Amar, and Franck Molina. Bionetcad : design, simulation and experimental validation of synthetic biochemical networks. *Bioinformatics*, 26(18) :2298–2304, 2010.
- [5] Patrick Amar, Guillaume Legent, Michel Thellier, Camille Ripoll, Gilles Bernot, Thomas Nystrom, Milton Saier Jr, and Vic Norris. A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. *BMC Systems Biology*, 2(27), 2008.