

Prédiction de la structure 3D des ARN liés à une protéine

La plupart des ARN exercent leur fonction dans la cellule en se liant avec des protéines, selon une géométrie 3D spécifique. Le **docking** consiste à prédire la structure 3D d'un assemblage moléculaire, à partir de la structure 3D de chaque molécule individuelle. Si la structure isolée d'une des molécules n'est pas connue expérimentalement, elle peut être modélisée. Dans le cas d'un ARN, de nombreux outils existent pour modéliser sa structure 3D à partir de sa séquence. Mais ces outils ne prennent pas en compte la liaison à une protéine. Or les ARN sont des molécules très flexibles, et leur structure 3D peut varier sous l'effet de leur liaison à une protéine.

En effet, la structure de plus basse énergie de l'ARN isolé ne correspond pas forcément à la structure de plus basse énergie du complexe ARN-protéine : une déformation de l'ARN coûteuse en énergie peut être compensée si elle permet une interaction de meilleure énergie de liaison avec la protéine. Cette flexibilité doit donc être prise en compte lors du docking, et si possible en amont lors de la modélisation de l'ARN isolé.

Pour cela, plusieurs options sont envisageables. D'une part, les méthodes de modélisation 3D des ARN isolés peuvent proposer un ensemble de solutions d'une taille choisie, ou jusqu'à un delta d'énergie choisi par rapport à la solution de plus basse énergie. Si cet ensemble est suffisamment grand, il devrait contenir la structure de l'ARN lié. D'autre part, il est possible de découper la structure d'un ARN en fragments, de docker ces fragments séparément sur la protéine, puis de réassembler les positions compatibles de fragments voisins [1]. Pour cela, il faut que chaque structure locale d'un fragment de l'ARN lié soit présente dans au moins une des structures prédites de l'ARN non lié.

Ces deux solutions seront envisagées successivement au cours du stage. La première étape consistera à tester quelle taille doit avoir un ensemble de structures 3D prédites d'ARN isolé pour avoir une forte probabilité de contenir une structure proche de celle de l'ARN lié. La deuxième étape consistera à évaluer la taille que doit avoir cet ensemble pour que chaque fragment de l'ARN lié (de taille à déterminer) ait sa structure présente dans au moins une des structures non-liées prédites. Nous utiliserons pour cela un ensemble de structures expérimentalement connues de complexes ARN-protéine. Des techniques algorithmiques faisant appel aux graphes, comme par exemple la recherche de cliques de tailles maximales, ainsi que l'optimisation multiobjectif (telles que utilisées dans [2] et [3] pour la prédiction de structures de complexes d'ARN (outils RCPred et C-RCPred disponibles sur la plateforme EvryRNA <http://EvryRNA.ibisc.univ-evry.fr>)), pourront être utilisées afin de déterminer les structures optimales de chaque fragment et la manière de les docker sur la protéine.

Si une de ces deux étapes aboutit aux résultats souhaités avant la fin du stage, nous simulerons l'amarrage des ensembles de structures 3D prédites sur le partenaire protéique, à l'aide de l'outil d'amarrage ATTRACT [4]. Le niveau de flexibilité à prendre en compte durant l'amarrage (par exemple par modes normaux et/ou par fragments) dépendra de la précision obtenue sur la structure de l'ARN.

Références :

[1] Chauvot de Beauchêne, De Vries, Zacharias. NAR 2016 ; Chauvot de Beauchêne, De Vries, Zacharias. [PLoS](#) 2016.

[2] Legendre, Angel, Tahi, RCPred: RNA complex prediction as a constrained maximum weight clique problem. [BMC Bioinformatics](#), 20-S(3): 53-62, 2019.

[3] Legendre, PhD thesis, Prédiction de structures secondaires d'ARN et de complexes d'ARN avec pseudonoeuds - Approches basées sur la programmation mathématique multi-objectif, Dec 2019

[4] Setny and Zacharias, NAR 2011