

SUJET DE STAGE M2 BIOINFORMATIQUE

Titre : Programmation parallèle et distribuée d'algorithmes bioinformatiques dédiés aux ARNs

Mots-clés : bioinformatique, structures d'ARNs, prédiction, Big Data, programmation distribuée et parallèle

Maître de stage : Fariza Tahy, Equipe AROBAS, laboratoire IBISC, UEVE/Genopole

Co-encadrement avec : Nazim Agoulmine, Equipe COSMO, laboratoire IBISC, UEVE/Genopole

Collaboration avec : Olivier Namy, Equipe « Génomique, Structure et Traduction » de l'I2BC (Institut de Biologie Intégrative de la Cellule) de Paris-Saclay.

Lieu du stage : Laboratoire IBISC, IBGBI, Université d'Evry, 23 Boulevard de France, 91000 Evry.

Durée de stage : 6 à 7 mois (à partir du 2 janvier)

Budget demandé : 3500 euros

Contact : Fariza Tahy : fariza.tahi@univ-evry.fr

Tel : 06 10 44 21 22

Résumé :

Les ARNs sont des molécules qui peuvent adopter de nombreuses structures secondaires. Ces repliements intra jouent un rôle majeur dans la fonction des ARN concernés. En plus de ces structures, certaines interactions en trans permettent aux ARNs de jouer des rôles régulateurs. Il a aussi été démontré qu'il était possible de modifier les propriétés traductionnelles des ARNs messagers au sein du ribosome grâce à l'hybridation d'ARN anti-sens créant des structures secondaires spécifiques.

La bioinformatique joue ici un rôle fondamental. Il s'agit dans un premier temps de prédire les structures possibles de complexes ARN anti-sens qui perturberaient le fonctionnement du ribosome, structures qui pourront ensuite être validées ou invalidées expérimentalement. Il existe d'une part de nombreuses méthodes de prédiction d'interactions intra ARN (notamment la structure secondaire), et d'autre part plusieurs méthodes de prédiction d'interactions inter ARNs de deux éléments. Or, à la manière des protéines et de leur structure quaternaire, des complexes plus importants d'ARNs se forment, où les interactions intra et inter moléculaires sont interdépendantes. D'où l'importance de prédire ces complexes.

Notre projet, étant donné la complexité du problème, est de développer des approches interactives, modulaires et distribuées pour prédire ces complexes moléculaires, et ce afin de proposer des outils fonctionnels, capables de passer à l'échelle.

Le but du stage est de développer la partie passage à l'échelle. Les méthodes de prédiction de structures et d'interactions d'ARNs existantes sont complexes en temps d'exécution et ne peuvent être utilisées que sur des séquences de tailles limitées. On adressera le problème de passage à l'échelle pour les différents algorithmes proposés, notamment ceux actuellement développés dans le cadre de la thèse d'Audrey Legendre, financée par l'ED STIC de Paris-Saclay (bourse établissement 2016). Cette thèse, qui s'inscrit dans le projet décrit ci-dessus, a pour but de proposer des algorithmes basés sur la programmation mathématique multi-objectif. Dans le projet, plusieurs autres approches algorithmes seront utilisées, notamment l'apprentissage automatique et l'algorithmique approchée. La programmation distribuée et la parallélisation de ces algorithmes seront utilisés pour prendre en compte les grands problèmes en utilisant les techniques de réduction appropriées. Un travail collaboratif avec d'autres membres du CDS, notamment autour des problématiques de Big Data, est envisagé.

Ce projet s'inscrit dans la thématique de recherche de Fariza Tahy qui développe depuis plusieurs années des algorithmes dédiés aux ARNs (<http://www.ncbi.nlm.nih.gov/pubmed/?term=fariza+tahi>). Tous ces algorithmes ont donné lieu à des logiciels disponibles sous forme de serveurs web sur la plateforme logicielle EvryRNA développée au sein de l'équipe (<http://evryrna.ibisc.univ-evry.fr>). EvryRNA est labellisée plateforme Genopole, et est largement utilisée au niveau international (autour de 2000 visiteurs par mois). Très récemment, Genopole, avec le soutien du conseil général de l'Essonne, a financé l'achat de plusieurs clusters et serveurs pour la plateforme, afin de permettre de meilleures performances en temps de calcul et donc un passage à l'échelle des différents logiciels fournis. Le stagiaire travaillera donc sur cette nouvelle infrastructure matérielle.