

## **Prédire les interactions protéine-protéine.**

Contact : [florian.richoux@univ-nantes.fr](mailto:florian.richoux@univ-nantes.fr)

### **I Contexte**

Les interactions protéine-protéine sont au cœur de tous les mécanismes de communication et de fonction biologique au sein d'une cellule. Comprendre et prédire comment les protéines interagissent entre elles est la clé pour comprendre les mécanismes du vivant au niveau cellulaire et ainsi mieux comprendre les maladies afin de proposer de nouveaux médicaments et de nouvelles méthodes thérapeutiques.

Il est estimé chez l'homme que de 100.000 à 600.000 interactions protéine-protéine différentes interviennent pour le bon fonctionnement des cellules. Or, nous ne connaissons seulement environ que 6.000 interactions protéine-protéine clairement identifiées au niveau atomique (c'est-à-dire, quels atomes de la protéine A interagissent avec quels atomes de la protéine B).

Il existe de nombreuses méthodes de prédiction d'interaction protéine-protéine basées sur des approches classiques d'apprentissage automatique, ou *machine learning*, mais elles n'ont jusque là permis d'étudier qu'une infime fraction de ces interactions.

Récemment, les méthodes d'apprentissage automatique suscitent un très fort intérêt de la part des communautés scientifiques et industrielles. En particulier, les méthodes dites d'apprentissage profond, ou *deep learning*, sont parmi les nouvelles techniques les plus populaires. Une particularité intéressante de ces méthodes réside dans le fait qu'elles isolent elles-même les caractéristiques qui lui semblent utiles dans les données, sans avoir besoin de pré-traiter ces dernières impliquant une intervention humaine dans la sélection de ces caractéristiques. Ceci est un avantage par rapport aux méthodes apprentissage automatique classiques, où les propriétés des objets étudiés sont extraites et sélectionnées par un humain qui les donne en entrée à l'algorithme d'apprentissage. Ceci induit donc un biais, parfois inconscient, en orientant l'algorithme d'apprentissage vers une certaine direction qui semble prometteuse.

### **II Stage de Master 2 (5/6 mois entre 01/18 et 08/18)**

L'objectif de ce stage est de proposer un format de données ainsi qu'un modèle de *deep learning* pour la prédiction d'interaction protéine-protéine. Ce projet de recherche se fait en étroite collaboration avec Stéphane Téletchéa du laboratoire de biologie "Unité Fonctionnalité et Ingénierie des Protéines" de l'Université de Nantes. Stéphane Téletchéa fournira les données brutes à traiter et apportera son expertise dans le domaine des interaction protéine-protéine.

Ce projet de recherche contient plusieurs étapes :

1. Dans un premier temps, ne prendre en compte que la chaîne polypeptidique (c'est-à-dire la chaîne d'acides aminés) composant chaque protéine. Le format de données sera ici simplement les deux chaînes polypeptidique des deux protéines à tester pour prédire si oui ou non elles interagissent entre elles. Il faudra proposer et expérimenter un modèle de *deep learning* apprenant ces interactions.
2. Avec les mêmes données, la seconde étape consiste à affiner la prédiction pour non plus déterminer si deux protéines interagissent, mais préciser à quel(s) niveau(x) dans leur chaîne polypeptidique respective l'interaction a lieu, s'il y en a une.
3. Enfin, l'objectif ultime de ce projet de recherche est de descendre au niveau atomique en prenant en compte la représentation 3D de la structure atomique de chaque protéine, allant bien au delà de la simple chaîne polypeptidique. Un véritable challenge se pose alors pour trouver un format de données convenable ainsi qu'un modèle de *deep learning* capable de prédire où se situe l'interaction au niveau atomique. Une telle méthode de prédiction serait inédite dans le domaine, ce qui souligne le caractère ambitieux de ce projet de recherche.

Ce stage exige plus des compétences de modélisation que de programmation de la part du/de la candidat-e. Pour la partie applicative du projet, la bibliothèque d'apprentissage profond *pytorch* sera utilisée. Les paramétrages expérimentaux se feront sur un serveur de calcul du Laboratoire des Sciences du Numérique de Nantes contenant plusieurs cartes GPU NVidia 1080 avant de passer en production sur le Centre de Calcul Intensif des Pays de la Loire sur une machine contenant plusieurs cartes GPU NVidia P100.

---

Ce stage rentre dans le cadre du projet interdisciplinaire de l'Université de Nantes.

Le stage se déroulera au Laboratoire des Sciences du Numérique de Nantes (LS2N) de l'Université de Nantes et sera gratifié au hauteur d'environ 550€ par mois (pour 22 jours par mois, 7h par jours). Il couvrira 5 à 6 mois sur la période janvier - août 2018.