

Développement d'algorithmes dédiés à la peptidomique

Guillaume FERTIN, guillaume.fertin@univ-nantes.fr

Géraldine JEAN, geraldine.jean@univ-nantes.fr

Hélène ROGNIAUX, helene.rogniaux@inrae.fr

- Date de début envisagée : 01/11/2024
- Établissement de conventionnement: Nantes Université
- Laboratoire de conventionnement: LS2N
- Financement : ANR PeptidOMS

La peptidomique fait référence à la branche de la protéomique dédiée à l'identification de fragments de protéines, appelés peptides, naturellement présents dans un système vivant. L'identification de tels peptides a des applications multiples et importantes, notamment en santé et en agronomie. Un défi majeur dans ce domaine consiste à fournir de nouvelles méthodes capables de détecter et d'identifier de manière efficace et fiable de tels peptides dans des matrices complexes telles que les fluides et tissus biologiques. Le suffixe « omics » implique des approches exploratoires à grande échelle, avec la production et l'exploitation de données analytiques massives. Comme pour d'autres sciences « omiques », la spectrométrie de masse tandem (MS/MS) s'est rapidement imposée comme la technique de base de la peptidomique et cette technologie est capable de traiter des échantillons d'une grande complexité et de produire des informations structurales de haute qualité sur des centaines de milliers de peptides en quelques heures.

Néanmoins, il n'est pas possible d'utiliser directement les méthodes classiques d'identification des peptides qui existent en protéomique sur des jeux de données issus de la peptidomique, car ces méthodes font l'hypothèse que les peptides sont tryptiques c'est à dire qu'ils ont été obtenus par digestion enzymatique avant d'entrer dans le spectromètre de masse. Sans cette hypothèse, l'espace de recherche augmente considérablement, et rend ainsi le problème considérablement plus difficile.

L'objectif de la thèse consiste donc à développer des nouvelles structures de données et des algorithmes spécialement conçus pour l'identification des peptides dans le domaine de la peptidomique prenant en compte le fait que ces peptides peuvent (i) commencer et terminer n'importe où dans une protéine et (ii) être fortement modifiés.

Dans un premier temps, l'objectif principal sera de permettre le passage à l'échelle en produisant un outil capable d'analyser des milliers de spectres contre un protéome entier en un temps raisonnable. Ensuite, l'objectif sera d'améliorer les prédictions en affinant le système de score à l'aide de techniques d'apprentissage automatique prenant en compte certains attributs physico-chimiques des peptides. Enfin, toujours dans le but d'améliorer à la fois le temps d'exécution et la pertinence des identifications, nous envisageons de comparer les spectres entre eux pour les regrouper et les traiter ensemble au sein d'un même cluster.

À chaque étape, les algorithmes développés seront soumis à des tests rigoureux sur des ensembles de données bien contrôlés et des données expérimentales qui seront fournies par les partenaires du projet PeptidOMS.

Compétences recherchées

Master en informatique ou en bioinformatique. Des connaissances sur les bases de données biologiques et leur manipulation n'est pas nécessaire mais sont appréciables.