

L'identification de protéines séquencées en mode MS/MS : des difficultés encore non résolues

Dominique Tessier

INRA Angers-Nantes

Cellule Bioinformatique et Gestion de données (BIOINF)

La question de l'identification par spectrométrie de masse de mélanges de protéines est un problème réputé complexe, encore aujourd'hui mal résolu, et qui reste un goulot d'étranglement dans le processus d'analyse. En spectrométrie de masse MS/MS, la fragmentation d'un peptide inconnu produit un spectre composé d'un ensemble de pics. L'interprétation de ce spectre par des méthodes dites *de novo* ou sa comparaison avec une banque de spectres de peptides connus, permet de proposer les séquences d'acides aminés les plus probables du peptide analysé. A partir de l'interprétation de plusieurs peptides issus d'une même protéine, il est ainsi possible d'identifier cette protéine. De nombreux algorithmes *de novo* sont disponibles, mais ils restent très sensibles à la très bonne qualité des spectres. Les algorithmes de comparaison des spectres sont, quant à eux, plus efficaces dès lors que le peptide analysé est bien présent dans la banque de référence, ce qui n'est pas le cas si le peptide analysé porte des modifications post-traductionnelles non connues à priori ou si l'espèce n'est pas séquencée.

La présentation fera un point sur les scénarii d'interprétation des spectres les plus usuels. Des études récentes pointent le taux encore faible d'interprétation dans ce contexte. Nous illustrerons des situations où l'interprétation reste délicate avec quelques exemples choisis pour montrer les limites des algorithmes actuels. Nous présenterons enfin les travaux réalisés dans la plate-forme analytique BIBS de l'INRA de Nantes : proposition d'une nouvelle modélisation d'un spectre MS/MS – Thèse de F. Cliquet -, élaboration de scénarii d'analyse – logiciel OVNIp -.