

La visualisation interactive d'information : une piste pour affronter la complexité des données d'identification par spectrométrie de masse ?

Renaud Blanch¹, Christophe Bruley²

1 : UJF–Grenoble 1/Laboratoire d'Informatique de Grenoble

2 : BGE, U1038 INSERM/CEA/UJF- Institut de Recherches en Technologies et Sciences pour le Vivant (iRTSV)

L'analyse par spectrométrie de masse d'échantillons biologiques parfois complexes passe souvent par une étape qui consiste à identifier les protéines présentes par l'analyse de fragments de ces protéines nommés peptides. Lorsqu'un peptide est analysé par le spectromètre de masse celui-ci enregistre un spectre de fragmentation qui va par la suite permettre de retrouver la séquence d'acides aminés qui compose le peptide. Si sur le papier ce processus semble simple, en pratique la détermination de la séquence peptidique correspondant au spectre de fragmentation est souvent entachée d'erreur et d'incertitude et l'appartenance d'un peptide à une protéine est parfois ambiguë puisqu'une séquence peptidique peut être retrouvée dans différentes protéines. La structure même d'un résultat d'identification est bien maîtrisée, et les scientifiques procèdent ensuite à une étape de validation qui tente de lever une partie de cette ambiguïté tout en contrôlant le taux de fausses identifications.

Cependant une analyse classique d'environ 60 minutes conduit à quelques dizaines de milliers de spectres et quelques milliers de peptides eux-mêmes identifiant un millier de protéines et faute d'outils adéquats pour explorer chaque analyse, les scientifiques prennent rarement le temps de regarder dans le détail ces résultats. En l'état actuel des solutions disponibles, lorsque les données sont explorées, elles le sont de manière statique, après la phase de validation, de manière exhaustive et répétitive ; et in-fine en masquant une partie de la complexité de la structure du résultat pour le rendre plus intelligible, mais au prix d'une perte d'information importante. La visualisation interactive d'information peut apporter des solutions à ce problème en proposant des représentations graphiques intelligibles des données avant qu'elles ne soient élaguées et agrégées. En concevant ces représentations graphiques pour tirer parti de la capacité du système visuel humain à détecter les régularités et les anomalies, celles-ci pourraient permettre, par exemple, de valider les résultats, de lever les ambiguïtés qui demeurent, de mettre en évidence les décisions basées sur des preuves de faible valeur, ou d'évaluer la profondeur d'analyse par comparaison à ce qui aurait pu/dû être identifié.

Au-delà de ce travail d'exploration et de qualification des données d'identification, l'exploration visuelle des données doit également s'intéresser à d'autres activités qui constituent le quotidien des experts protéomiciens : il est ainsi fréquent de s'intéresser aux similitudes et aux différences qui peuvent exister entre des échantillons biologiques ou entre des collections d'échantillons, soit sur la base de la seule présence ou absence des protéines identifiées soit en utilisant des méthodes qui vont permettre de quantifier de façon plus ou moins fine ces différences. Là encore, l'analyse visuelle des données de protéomique quantitative est un enjeu d'actualité.