

# Analyse statistique des performances des méthodes de quantification protéique

Klich A<sup>1</sup>, Mercier C<sup>1</sup>, Grangeat P<sup>2</sup>, Gerfault L<sup>2</sup>, Charrier J-P<sup>3</sup>, Maucort-Boulch D<sup>1</sup> et Roy P<sup>1</sup>

<sup>1</sup>Hospices Civils de Lyon, Service de Biostatistique; Université Lyon 1 ; CNRS UMR 5558, Laboratoire Biométrie et Biologie Evolutive, Equipe Biostatistique Santé, Lyon, France.

Plateforme Rhône-Alpes de BioInformatique à haut débit (PRABI).

<sup>2</sup>CEA-Leti, MINATEC Campus, Health Division

<sup>3</sup>bioMérieux, Technology Research, Mass Spectrometry Platform

[amna.klich@chu-lyon.fr](mailto:amna.klich@chu-lyon.fr), [catherine.mercier@chu-lyon.fr](mailto:catherine.mercier@chu-lyon.fr),  
[pierre.grangeat@cea.fr](mailto:pierre.grangeat@cea.fr), [laurent.gerfault@cea.fr](mailto:laurent.gerfault@cea.fr), [jean-philippe.charrier@eu.biomerieux.com](mailto:jean-philippe.charrier@eu.biomerieux.com), [delphine.maucort-boulch@chu-lyon.fr](mailto:delphine.maucort-boulch@chu-lyon.fr), [pascal.roy@chu-lyon.fr](mailto:pascal.roy@chu-lyon.fr)

## Abstract

Pour étudier et comparer les performances de trois méthodes de quantification de la concentration protéique, nous proposons un plan expérimental et un plan d'analyse prenant en compte les différentes sources de variabilité dans la chaîne d'analyse par *spectrométrie de masse en mode SRM*.

# 1 Introduction

Les techniques de spectrométrie de masse sont très séduisantes pour détecter des biomarqueurs protéiques de manière sensible et rapide. Cependant, contrôler la variabilité technique sur ces chaînes d'analyse reste un enjeu majeur pour la recherche et la validation des biomarqueurs. Dans le cas de la chaîne d'analyse par spectrométrie de masse en mode SRM, une technique de traitement bayésienne de l'information intégrant la variabilité technique a été proposée sur le projet BHI-PRO, dans le contexte de la validation de biomarqueurs. Un algorithme d'inversion hiérarchique bayésienne estime conjointement les paramètres biologiques et techniques inconnus et permet de calculer la concentration protéique. Cet algorithme a pour objectif de diminuer la variance de l'estimation due à la variabilité technique. Pour évaluer correctement cette part de variance, il est nécessaire de quantifier la variabilité technique attribuable aux diverses étapes de la chaîne analytique et de la distinguer de la variabilité biologique. La difficulté est que la variabilité biologique n'est pas connue, d'une part en raison des variabilités individuelles entre patients et d'autre part parce qu'il n'existe pas de méthode de référence parfaite pour estimer la concentration protéique. La méthode ELISA, souvent considérée comme une référence, est elle-même imparfaite et ne permet pas d'estimer la concentration sans erreur. Une manière de maîtriser la variance biologique est de la créer artificiellement par dilution (rampe de dilution). La concentration relative varie alors selon le facteur de dilution sans que l'on connaisse la concentration absolue. Nos objectifs sont :

- 1) Proposer un plan expérimental pour quantifier la variabilité technique attribuable aux diverses étapes de la chaîne analytique sur des données artificielles en rampe de dilution.
- 2) Comparer les résultats fournis par deux algorithmes de traitement des données SRM aux résultats de la méthode ELISA. Les performances de l'algorithme d'inversion hiérarchique bayésienne seront comparées à celles d'un algorithme usuel.

## 2 Matériel et méthodes

Les étapes de la chaîne analytique en mode SRM sont nombreuses et incluent principalement la préparation des échantillons, la digestion des protéines et le fractionnement des peptides, la séparation des peptides par chromatographie liquide, leur injection dans le spectromètre de masse par électro-ionisation, la sélection et le fractionnement des peptides ciblés, et la détection des fragments d'ions. La variabilité de la mesure résultant de ces étapes se situe au niveau du passage entre le niveau protéine et le niveau fragment qui fournit le signal, appelé transition. Les composantes discernables de cette variabilité sont celles du passage protéine-peptide après fractionnement et celle du passage peptide-transition. La variabilité due au passage protéine-peptide provient de la préparation et de la digestion. Elle se traduit notamment par une variabilité du gain correspondant. Ce gain est estimé chaque jour par calibrage externe avec un mélange étalon. Ceci permet d'introduire dans l'algorithme de traitement du signal une correction. Mais cette correction reste entachée d'une variabilité. La variabilité due au passage peptide-transition provient essentiellement de l'étape LC-MS, le gain de ce passage étant estimé par calibrage interne pour chaque mesure avec des peptides alourdis, selon l'approche AQUA. La variabilité de la mesure comprend aussi la variabilité due à l'algorithme utilisé selon sa prise en compte des fluctuations sur la mesure et selon les estimateurs utilisés. L'estimation de la concentration protéique par l'algorithme usuel est calculée directement à partir de transitions sélectionnées. Elle repose sur une détection et une quantification des pics sur les traces SRM. L'estimation de la concentration protéique par l'algorithme bayésien utilise toute la hiérarchie de quantification (protéine-peptide-fragment) pour fusionner les informations délivrées par chaque trace

associée à une transition ciblée de la protéine étudiée. L'algorithme bayésien repose sur un modèle paramétrique intégrant les principales causes de variabilité comme par exemple les temps de rétention chromatographiques. La méthode d'inversion estime alors conjointement les paramètres du modèle et les concentrations des protéines.

Pour acquérir les données de la rampe de dilution, nous proposons un plan expérimental qui prend en compte les contraintes de laboratoire et permet d'estimer les variances inter-jour et intra-jour induites par les passages protéine-peptide et peptide-transition. Les données seront obtenues à partir d'échantillons d'un mélange étalon utilisé pour le Contrôle Qualité de la chaîne de mesure. Ce mélange étalon est un sérum sanguin « poolé » contenant des protéines diluées, notamment les protéines ciblées. Ce mélange étalon est ensuite dilué afin de produire une gamme de dilution. Les facteurs de dilution appliqués valent respectivement : 1, 1/2, 1/4, 1/8, et 1/16. Cette série d'échantillons est mesurée conjointement par la technique de spectrométrie de masse SRM et par une technique ELISA réalisant le dosage de protéine par anticorps. Dans cette communication, nous proposons de décrire les méthodes d'analyse statistique que nous souhaitons appliquer pour :

- 1) Vérifier la linéarité de la relation entre la concentration estimée et le facteur de dilution par comparaison de modèles emboîtés à effets mixtes;
- 2) Décomposer la variance des estimations des concentrations des analytes pour les différentes méthodes d'estimation (par chacun des 2 algorithmes SRM ou par ELISA).

### 3 Résultats escomptés

Le plan expérimental permettra d'évaluer les contributions respectives des variabilités biologique et technique à la variabilité de la mesure. L'objectif est d'étudier l'amélioration apportée par le nouvel algorithme bayésien qui devrait permettre de diminuer le rapport variance technique/variance de la mesure. La méthode ELISA devrait permettre de mieux examiner et comparer les performances des méthodes SRM.

### Références

Brown, H. and Prescott, R. (2006). *Applied mixed models in medicine*. 2<sup>nd</sup> Ed. Wiley.

Gerfault L., Huillet C., Brun V., Grangeat P. (2011), "Automatic and adaptive processing algorithm of SRM traces for protein quantitation", Human Proteome Organization (HUPO) 10<sup>th</sup> annual world congress, 4-7 September 2011, Geneva, Switzerland.

Gerfault L., Szacherski P., Giovannelli J.-F., Charrier J.-P., Mahé P., Grangeat P. (2012), "A hierarchical SRM acquisition chain model for improved protein quantification in serum samples", Research in Computational Molecular Biology (RECOMB) Satellite Conference On Computational Proteomics 2012, San Diego, USA, 6-8 avril 2012.