A model for the analysis of proteolytic ^{18}O stable-isotope labeled peptides in MALDI-TOF mass-spectra.

Dirk Valkenborg, Tomasz Burzykowski

Center for Statistics, Hasselt University, Agoralaan - building D, 3590 Diepenbeek, Belgium

The intention of differential proteomic is to compare distinct proteomes between two or more conditions to obtain a set of proteins, which can be differentially expressed beween these conditions. Several experimental techniques are available for this purpose. Typically, they involve electrophoresis and/or chromatography for separating the complex biological samples and mass spectrometry for analyzing the content of the samples. However, the random noise introduced by chromatography and mass spectrometry makes it difficult to compare the proteomic profiles coming from the groups of samples. Therefore, stable-isotope coding is often used, so that peptides from distinct groups can be pooled together and analyzed simultaneously on the mass spectrometer. In other words, protein information about different conditions appear together in a single mass spectrum, making a direct comparison possible (in terms of, e.g, the ratio of measured total intensities).

A powerful and relatively new technique for stable-isotope coding is the proteolytic ^{18}O labeling. In this setting, oxygen atoms from the carboxyl-terminus (COO^{-}) of peptides are replaced with oxygens from heavy-oxygen-water during proteolysis. However, variable ^{18}O incorporation rate, ^{16}O back-exchange, and impurities (presence of ^{16}O and ^{17}O atoms) in the heavy-oxygen-water result in multiple isotopic distributions from the labeled peptide superimposing with the unlabeled peptide. Thus, it is by no means an easy task to determine the exact intensity ratio from the observed spectrum.

We propose a model, similar in spirit to that developed by Eckel-Passow *et al.* [1], which estimates the isotopic distribution, peptide-specific incorporation rate, and the intensity ratio directly from the observed mixture of peptide peaks. The model requires the knowledge of the composition of the heavy-oxygen water. We include a possibility of label-swap (cfr. dye-swap with cDNA microarrays) and argue that it can improve the identifiability and estimability of the model. The performance of the method is illustrated using real-life datasets.

[1] Eckel-Passow J, Oberg A, Therneau T, et al. Regression analysis for comparing protein samples with 16O/18O stable-isotope labeled mass-spectrometry. *Bioinformatics* 2007; **2**:305-318.