6th International Symposium on
RECENT ADVANCES IN FOOD ANALYSIS
November 5 - 8, 2013 • Prague, Czech Republic
RAFA 2013

Tutorial on

tools for mass spectrometry-based metabolomic data processing and analysis

Moderator:

Dr. Tomas Cajka
(Metabolomics Fiehn Lab, Genome Center, University of California, Davis, USA)

Thursday, 7 November 2013, 16:00 – 18:00 (Leo & Virgo halls)

The metabolomics approach, aiming at global analysis of numerous targeted or non-targeted low molecular compounds (metabolites) in a biological sample, has recently found its application in diverse research areas including food quality and safety, and authenticity. A rapid growth of metabolomics has been enabled by substantial advances in analytical techniques such as mass spectrometry (MS) coupled to liquid chromatography (LC) or gas chromatography (GC), and nuclear magnetic resonance (NMR), all the techniques facilitating analysis of a wide range of metabolites with diverse physicochemical properties and occurring at different concentration levels. To process and interpret the complex data obtained within metabolomic-based studies, advanced software algorithms of data handling are needed, consisting of data processing, data pretreatment, and data analysis.

Data processing proceeds through multiple stages such as filtering, peak detection, deconvolution, alignment, and normalization. The need of powerful data-processing methods gave rise to numerous commercial as well as free tools implementing one or several steps of the data processing pipeline.

Data pretreatment represents another crucial step that can dramatically change the outcome of the data analysis. This procedure typically involves centering and scaling of the original data to eliminate unwanted systematic bias, while maintaining genuine differences in the examined datasets.

Data analysis involves the use of various chemometric tools. Unsupervised pattern recognition techniques (represented mainly by principal component analysis) are often the first step of the data analysis in order to detect patterns in the measured data. On the other hand, supervised pattern recognition techniques (*e.g.*, partial least-squares discriminant analysis, linear discriminant analysis) use the existing information about the class membership of samples to a given group (class or category) to classify a new "unknown" sample using its pattern of measurement. From this point of view, the outputs of metabolomic data analysis may differ depending on the purpose of investigation.

In most cases, there is also an interest in ***identification*** of the discriminating (bio)marker compounds. In this case, the use of a high-resolution instrument enabling to obtain both single MS and MS/MS accurate mass spectra is needed for reliable elemental formula estimation, which is typically followed by a database search.

In this tutorial, a brief overview of the key steps involved in data handling and interpretation in mass spectrometry-based metabolomics (covering GC–MS and LC–MS instrumental platforms) will be presented. Case studies and practical examples will be used to demonstrate these concepts.