GC×GC-(HR)-TOFMS for Metabolomics and Volatilomics

J.-F. Focant[1], D. Zanella[1], N. Di Giovanni[1], M.-A. Meuwis[2], E. Louis[2], F. Schleich[3], R. Louis[3], P.-H. Stefanuto[1]

[1]Organic and Biological Analytical Chemistry Group, University of Liège, Belgium
[2]Gastroenterology Medicine, GIGA I3, CHU Sart-Tilman, Belgium
[3]Respiratory Medicine, GIGA I3, CHU Sart-Tilman, Belgium

Summary: *GC×GC-(HR)-TOFMS is a powerful hyphenated technique for the consideration of (semi)-volatile analytes. It is used for the search of biomarkers of illness for better understanding of metabolic processes in medical applications. This includes profiling VOC signatures in breath, in the headspace of bio-fluids and cultured cell media, or performing direct metabolite screening in fluids.*

Keywords: *Chromatography; Mass Spectrometry; Data Mining.*

Introduction

Comprehensive two-dimensional gas chromatography (GC×GC), especially when coupled to time-of-flight mass spectrometry (TOFMS), has come a long way since the original report on the technique in early 1990's. Several robust modulation devices (cryogenic modulators, flow modulators, solid state modulators, …) are now available and cover most fields of applications. Current GC×GC-TOFMS instruments allow to separate thousands of signals for semi-routine analyses of large sets of samples. Data processing requires complex dedicated procedures to properly extract the relevant information from the tremendous quantity of multidimensional data, especially when high resolution (HR)TOFMS is used.

Amongst the numerous fields of application of GC×GC-TOFMS, medical applications are of high interest as they ultimately can play a role in public health at a large scale. Such applications include search for biomarkers of illness and understanding of metabolic processes. These type of studies usually require profiling volatile organic compounds (VOCs) present in breath; measuring headspace VOC signatures from blood, urine, feces, and cultured cell media; or performing metabolite screening on plasma samples…

For each of these, it is important to develop and optimize proper sampling procedures often based on solid-phase microextraction (SPME) or thermal desorption (TD). The implementation of robust chromatographic methods is also crucial to ensure that high quality signals are produced prior to any data processing. Then, the most important aspects to be considered are data alignment, data reduction, (un)supervised statistics, and visualization of results. Several data mining approaches and statistical tools (Fisher ratio, PCA, PLS, clustering, machine learning…) have been implemented to digest the large amount of data generated. This will be illustrated through recent achievements in the medical field in terms of volatilomics and metabolomics.

Methods

We developed and implemented several GC×GC-HRTOFMS methods for both target and untarget metabolomics as well as volatile profiling of human samples such as exhaled breath, serum, and bronchoalveolar lavage fluids (BALF) under strict QA/QC guidelines. We developed a robust data processing approach based on templates relying on GC and MS quality features, and multiple (un)supervised statistical techniques.

Results

Breath analyses were carried out in several volatilomic studies on lung cancer and asthma. In both cases, GC×GC-HRTOFMS permitted to isolate putative biomarkers of disease that were selected

during model creation of test sets further applied on training set for testing of the robustness. In asthma phenotype studies over more than 150 patients, sets of markers were validated for the differentiation of phenotypes, as illustrated in ROC curves displayed in Fig. 1.

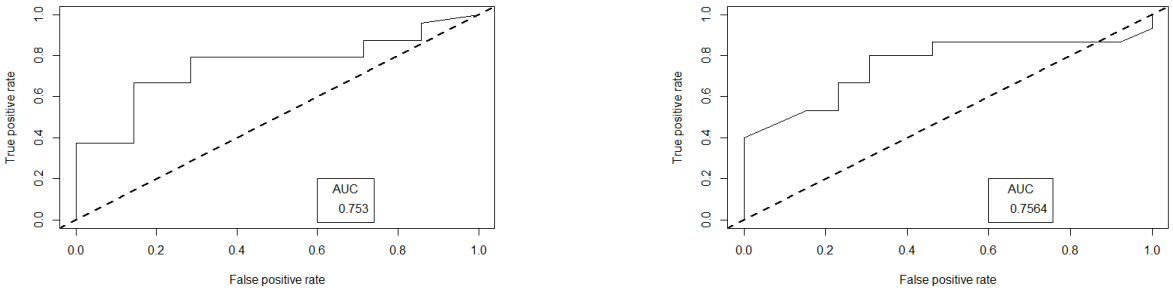

Fig. 1 : ROC curves for differentiation of neutrophilic and Paucigranulocytic (left) and eosinophilic and Paucigranulocytic (right) phenotypes based on a set of respectively 2 and 3 selected markers (exact identities known but not under appropriate protection for diffusion).

The use of Fisher ratio (FR) calculation using a statistical cut off value or classical Random decision forests allowed to extract analytes of interest from the large matrices of features produced by GC×GC-HRTOFMS. The use of MS libraries, the high mass accuracy (<1ppm) of the MS analyzers, and the linear retention indices (LRIs) allowed to known the exact identities of most isolated markers. PCA and HCA were used for data reduction and visualization purposes. Fig. 2 illustrates the PCA and HCA for a study on inflammatory bowel disease (Crohn's disease) for a class of blood samples (30 µL) from highly inflamed patients versus a class of healthy controls.

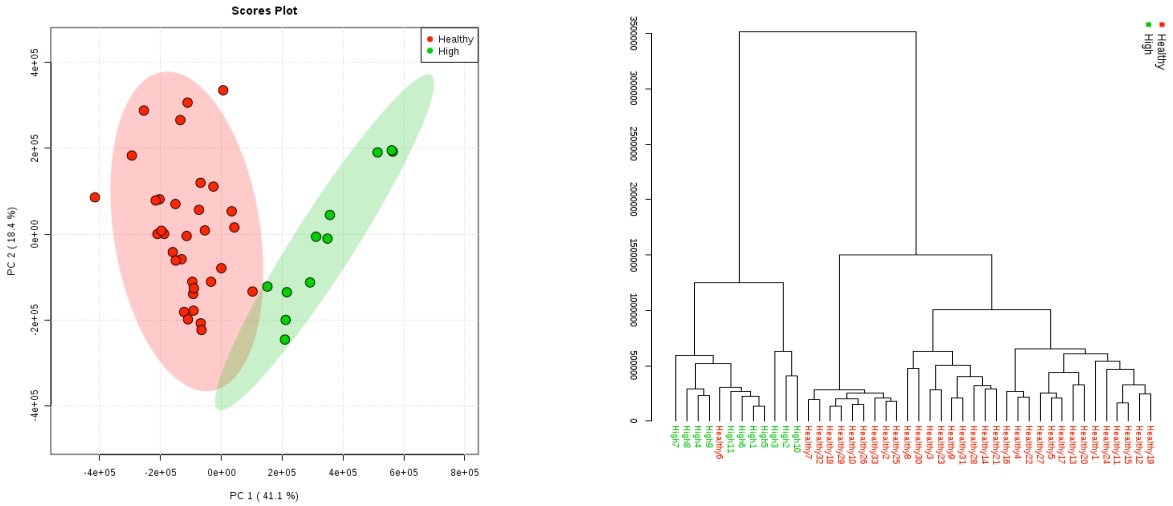

Fig. 2 : PCA score plot (left) and HCA plot (right) for the separation of Crohn's disease patients from controls based on a set of 14 putative biomarkers isolated from blood samples (exact identities known but not under appropriate protection for diffusion).

System stability was ensured based on the monitoring of quality control charts over time for retention time deviation in both first and second dimension, as well as instrumental relative response against selected standards. Quality control samples and NIST SRM were used and LOESS regression used to ensure reliability of the data.

Conclusions

GC×GC-HRTOFMS and appropriate data reduction/treatment is a valuable approach for serum metabolomics and volatilomics in the context of medical applications. High levels of QA/QC are requested to offer the requested level of robustness and possible future clinical use.