Gas chromatography mass spectrometry: key technology in metabolomics
Koek, Maud Marijtje

Citation

Version: Corrected Publisher’s Version
License: Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from: https://hdl.handle.net/1887/14328

Note: To cite this publication please use the final published version (if applicable).
Conclusions and perspectives

In a true systems biology approach, investigating an organism or biological system as an integrated system of interacting genes, proteins and metabolites, coupling of data on transcriptomics (gene expression), proteomics (protein analysis) and metabolomics is key. To explain the functioning and response to perturbations of, for example a human being in relation to health and disease, one would like to construct models, consisting of networks of interrelated genes, proteins and metabolites throughout the entire body. Ultimately, a successful systems biology approach will open new and exciting ways to improve the quality of human life based on individual phenotypes, for example through personalized health strategies, including medicine, diet and life style.

Metabolomics involves the unbiased quantitative analysis of all metabolites in cells, tissues and body fluids. Gas chromatography (GC) coupled to mass spectrometry (MS) is a very suitable technology for metabolomics, as it combines high separation efficiencies with generic, sensitive and selective mass detection. In addition, the elaborate fragmentation obtained with electron ionization facilitates the assignment of already identified metabolites and ultimately the de novo identification of so far unidentified metabolites.

In this thesis, the optimization, validation and application of GC-MS and comprehensive two-dimensional gas chromatography-mass spectrometry (GC×GC-MS) is described, relying on derivatization via oximation and subsequent silylation. Throughout the research emphasis was put on obtaining reliable and quantitative analytical data. In principle, GC×GC-MS is preferred to GC-MS in metabolomics analysis as (i) higher peak capacities lead to cleaner mass spectra, (ii) the detectability of metabolites is improved and (iii) lower detection limits are obtained. In addition, by using a non-conventional column setup in GC×GC-MS, i.e. a polar x apolar combination with a wider-bore and thicker-film second dimension column, an improved quantification of low-abundant metabolites in the presence of high abundant matrix compounds is achieved compared to conventionally used GC×GC-MS setups and GC-MS (Chapter 5). However, non-target quantification of all metabolites in GC×GC-MS chromatograms is still a major challenge. To get from the raw data to a list of metabolites and their corresponding peak areas (corresponding to concentrations), the peaks need to be correctly deconvoluted, integrated and the peaks from different modulations originating from one compound need to be merged. The available software for GC×GC-MS data processing is capable of providing deconvoluted peak lists. However, mistakes in the deconvolution and merging of peaks
lead to less precise quantification compared to a targeted approach. Nevertheless, the added value of GC×GC-MS is demonstrated in Chapter 6, in preclinical study on the development of diabetes type 2 in mice. Although the RSDs obtained with GC×GC-MS were higher than with GC-MS due to the data processing, the biological information acquired in GC-MS was preserved and several extra candidate biomarkers for the development of insulin resistance were found in the GC×GC-MS data set. However, the speed of the software algorithms limit the application of automated non-target quantification to small studies, with no more than approximately 30 – 50 samples (eventually measured in duplicate). Further improvement of data-processing software tools is required to apply GC×GC-MS to larger data sets and to fully benefit from the advantages of GC×GC-MS over GC-MS.

The most interesting and important technological developments for future research in metabolomics analysis in general are related to metabolite identification, improving the coverage of the metabolome, development of improved data-handling tools and improved throughput. In addition, the validation of the biological information obtained from metabolomics studies is of utmost importance to demonstrate the true added value of the non-targeted comprehensive or holistic systems biology approach. Metabolite identification in metabolomics analysis is a very complicated task due to the complexity of typical biological samples. To obtain relevant biological information from the acquired data and ultimately to answer the biological questions underlying metabolomics studies, knowing the identity of measured peaks is essential. However, still a large percentage of metabolites measured in metabolomics analysis remain unidentified. Many academic groups, contract research organizations and commercial companies are presently working on the improvement of metabolite identification, either by injection of as many as possible reference standards, building databases with retention times and mass spectra of metabolites and/or developing better routines and strategies for identification of unknown metabolites in biological samples. For GC-MS related techniques, the mining of (electron ionization) mass spectra for information on metabolite identity is promising, for example mass spectra can be used to find common fragments or losses and relate these to functional groups present in a molecule. Such data-mining tools are already available in proteomics analysis and also in GC-MS interpretation software (e.g., NIST, MS interpreter), and the applicability of such mining tools in GC-MS based metabolomics analysis should be investigated and eventually optimized and improved. In addition, in GC×GC-MS the position of a metabolite in the chromatogram can reveal information on its identity. As can be seen in Figure 1, order was found in the elution of structurally related compounds. Therefore, the use of retention indices in both the first and the second dimension could be helpful for identification purposes. Furthermore, the coupling of GC-based techniques to high-resolution mass spectrometry, possibly using atmospheric-pressure ionization or chemical ionization, or ion trap mass spectrometry, Fourier transform
Conclusions and future perspectives  Chapter 8

Infrared spectroscopy (FTIR)\(^7\) or nuclear magnetic resonance (NMR)\(^8\) can provide further insight in the identity of metabolites. However, the detection limits of NMR and FTIR are much higher than those of mass-spectrometry-based detection and the identification of metabolites with capillary GC-FTIR and GC-NMR is limited to high concentrated metabolites.

Figure 1 Ordering of structurally related metabolites in the two dimensional plane of a GC\(\times\)GC-MS chromatogram (polar \(\times\) apolar column combination).

The ultimate goal in metabolomics is to obtain reliable quantitative data for all metabolites present in body fluids, cells or tissues. This is a challenging task for analytical chemists, as metabolites originate from many different compound classes and, moreover, the concentration ranges of metabolites within a biological sample are extremely large (> 9 decades). To achieve this goal a combination of different analytical methods will be required. For example, LC-MS methods are quite complementary to GC-MS methods, as the metabolites that are not suitable for GC based analysis, for example due to their high molecular weight or ionic properties, are generally very suitable for LC-MS analysis. However, the accurate quantification of trace metabolites, such as signal metabolites or hormones, is so far virtually impossible using a non-targeted comprehensive approach such as the developed GC-MS and GC\(\times\)GC-MS methods; it will require selective analytical methods with lower detection limits instead. For example, sensitivity could be increased by a selective and targeted GC-MS/MS method or selective derivatization. Furthermore, the application of more selective methods could also aid in the identification of metabolites found in comprehensive methods. Finally, the extraction of the relevant biological information from the separate analytical techniques requires the data sets to be merged. Not only
merging tools, but, more importantly, protocols for data merging need to be developed to be able to explore differences in (relative) metabolite concentrations acquired with different analytical methods. Obtaining absolute metabolite concentrations rather than relative concentrations will help make data comparable and will help to integrate data of different levels (metabolites, proteins, transcripts, SNPs, etc.) often acquired by different groups.

In the last years, comprehensive GC-MS (Chapter 3) has become a routinely applied method in our lab. The method facilitates the reliable quantification of hundreds of different metabolites and has been successfully applied in many different studies in the fields of microbial and mammalian metabolomics, of which many have been requested and funded by industry. Important trends for future developments in comprehensive GC-MS analysis will be (i) faster and fully automated analysis, (ii) better coverage of the metabolome and (iii) miniaturization of the sample workup. With the use of smaller bore columns in GC-MS faster analysis and lower detection limits can be achieved. The application of GC×GC-MS will substantially contribute to improving the coverage of the metabolome and to the detection and quantification of low-abundant metabolites. For now, the lack of optimal data-processing tools limits the use of GC×GC-MS to small studies. However, the routine application of GC×GC-MS probably will become feasible in the next years. Of course, the identification of metabolites is of key importance to be able to determine the coverage of metabolomics methods. Miniaturization of the sample workup is required to apply GC-MS based metabolomics in very small volumes of sample, for example in mammalian studies with small animals with only a few µL of available sample or, even more demanding, in single-cell analysis with sample volumes in the nL range or below. The feasibility of comprehensive GC-MS in an aliquot (about 100 nL) of intracellular content of a single *Xenopus laevis* oocyte (with an internal volume of about 1 µL) and an aliquot of only ten nL of serum have been demonstrated in this thesis (Chapter 7). The possibilities of this method need to be further explored; it is expected to provide more inside into local biochemical processes in animals or ultimately in individual cells.

In conclusion, as of now, the performance and comprehensiveness of GC-MS and especially GC×GC-MS in quantitative comprehensive metabolomics analysis is unequalled by any other routinely applied technique (this thesis). Therefore, GC-MS has become a key technology in metabolomics analysis for systems-biology studies and will remain an essential asset in the metabolomics toolbox.
Reference List
