General introduction and scope

**Metabolomics**

The importance of measuring metabolites in body fluids has been long known. Already in ancient time (1500 – 2000 BC), Chinese doctors used ants to detect glucose in urine of patients to diagnose diabetes, and more related to present time, in the clinical-chemistry field gas chromatography has been used for decades to profile metabolites in body fluids for the diagnosis of diseases. Metabolomics evolved from the conventional profiling techniques combined with the idea to study organisms or biological systems as an integrated system of interacting metabolites, proteins and/or genes in pathways and cellular processes, the so-called systems biology approach. In analogy with other functional genomics technologies, such as transcriptomics (global analysis of gene expression) and proteomics (measure the complete proteome), metabolomics involves the unbiased quantitative and qualitative analysis of the complete set of metabolites present in cells, body fluids and tissues (the metabolome).

The metabolome is closely related to the genotype and physiology of an organism as well as environmental conditions. The biochemical level of the metabolome is closest to that of the function of a cell (the phenotype), and therefore, the study of the metabolome is key in understanding biological functioning. Ultimately the goal in metabolomics analysis is to obtain quantitative data and identifications for all metabolites in samples in order to find answers to a biological question and/or to relate metabolites to desired or undesired phenotypic features. A schematic of a typical workflow in metabolomics studies is shown in Figure 1. For a metabolomics study to be successful, all steps, starting from the definition of the biological question and the experimental design up to the biostatistics, should be optimized and appropriate for their intended use. In the final step metabolites relevant to a specific phenotypic characteristic can be identified. To extract the relevant information from the very large data set obtained from the analysis of all samples usually requires the use of multivariate data analysis (MVA) and/or pattern-recognition tools, such as principle-component analysis (PCA), principle-discriminant analysis (PCDA), partial least square (PLS), etc. An overview of frequently used MVA tools for metabolomics is provided by van der Greef en Smilde.
**Metabolome**

The metabolome refers to the complete set of metabolites present in a biological sample. Metabolites are the products and intermediates of metabolism and are usually restricted to small organic molecules (Mw < 1000). Metabolites are interrelated in a metabolic network, consisting of many different metabolic pathways. Part of the human metabolic network and an example of a metabolic pathway are shown in Figure 2. The complete set of metabolic pathways is important for maintaining a (healthy) constant stable environment within an organism, so-called homeostasis.
Metabolites originate from many different compound classes, e.g. alcohols, aldehydes, amines, amino acids, aromatic compounds, carbohydrates (sugars, sugar-phosphates, sugar-amines, sugar acids, etc), lipids (fatty acids, mono-, di- and triglycerides, phospholipids, sphingolipids, etc.), nucleosides (purines, pyrimidines), nucleotides and organic acids (carboxylic acids, sulfonic acids, phosho-organic acids, etc.). The complexity of the full metabolome is dependent on the organism studied, varying from a few hundreds of endogenous metabolites for microorganisms to thousands of endogenous human metabolites (The Human Metabolome Database already contains over 6500 distinct metabolites and the number is growing). In addition, by estimation hundreds of thousands of small molecules can be expected to be present in humans due to the microbes present in their guts or the administration of food, drugs, etc.

Applications
Metabolomics is increasingly used in a variety of applications in microbiology, plant- and medical science (pharmacology, clinical chemistry). For example, metabolomics can be used to improve the production of a (bio)chemical in a microbial production process; by determining bottlenecks in the production process, or improving the environmental conditions for growth. An example of such an approach is illustrated in Figure 3. The production of vitamin B12 by Propionibacterium freudenreichii was studied under different environmental conditions (different growth media) and at different time points during growth. After analysis by gas chromatography and liquid chromatography mass spectrometry the data were analyzed using PLS, i.e. identifying the metabolites correlating the most (positively or negatively) with vitamin B12 production (Figure 3). Subsequently, this ranked list of potential bottlenecks was biologically interpreted. As an example, the glucose-6-phosphate (glu-6-P)/fructose-6-phosphate (fruc-6-P) metabolite pair was present in this top 12 (Fig. 3). These two metabolites are part of the glycolysis and interconverted by the enzyme glucose-6-phosphate isomerase. In principle, all enzymatic reactions present in a microbial cell catalyze equilibrium reactions and the substrate and product concentrations of any enzyme should abide the rules of thermodynamics. In this study, a discrepancy between the thermodynamic equilibrium and the actual ratio between glu-6-P and fruc-6-P was found from the measurements, indicating a potential point of metabolic regulation or bottleneck in the vitamin B12 biosynthesis route. The P. freudenreichii strain was subsequently improved by overexpressing the glucose-6-phosphate isomerase gene, resulting in a 17% increase in vitamin B12 production by this industrial production strain. Furthermore, the addition of extra glycine, another metabolite strongly correlating with vitamin B12 production, to the growth medium resulting in an 12% increase in vitamin B12 production (unpublished results, courtesy of Mariët van der Werf, TNO). In analogy, metabolomics can be used to manipulate plant species.
Another frequently used application of metabolomics involves the study of diseases in clinical studies or studies using animal models. By comparing the metabolomes of healthy and diseased subjects, possible biomarkers for disease can be identified, or new insights can be gained in the development or progression of a disease. An interesting example of such an application was recently published by Sreekumar et al. They used LC-MS and GC-MS based metabolomics to investigate the development and progression of prostate cancer. By profiling the metabolomic alterations of prostate cancer progression, they revealed sarcosine as a potentially important metabolic intermediary of cancer cell invasion and aggressivity.

![Thermodynamic equilibrium K_{eq}](image)

<table>
<thead>
<tr>
<th>Compound</th>
<th>Regression value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Trisaccharide</td>
<td>0.0493</td>
</tr>
<tr>
<td>2  α-ketoglutaric acid</td>
<td>0.0486</td>
</tr>
<tr>
<td>3  Malic acid</td>
<td>0.0419</td>
</tr>
<tr>
<td>4  3-Phosphoglyceric acid</td>
<td>0.0401</td>
</tr>
<tr>
<td>5  Alanine</td>
<td>0.0387</td>
</tr>
<tr>
<td>6  Glyceric acid</td>
<td>0.0386</td>
</tr>
<tr>
<td>7  Fumaric acid</td>
<td>0.0363</td>
</tr>
<tr>
<td>8  Glucose-6-phosphate</td>
<td>0.0344</td>
</tr>
<tr>
<td>9  Fructose-6-phosphate</td>
<td>0.0344</td>
</tr>
<tr>
<td>10 Sedo-heptulose-7-phosphate</td>
<td>0.0343</td>
</tr>
<tr>
<td>11 Aspartic acid</td>
<td>0.0329</td>
</tr>
<tr>
<td>12 Glycine</td>
<td>0.0325</td>
</tr>
</tbody>
</table>

Figure 3 Identifying bottlenecks in microbial metabolomics by correlation with thermodynamic data.

**Analytical techniques**

The development of generic methodologies to analyze the complete metabolome, or at least as many metabolites as possible, is very challenging considering the complexity of the metabolome. Moreover, metabolites originate from a wide variety of compound classes with different physicochemical properties and are present in a large range of concentrations (> 9 decades). First of all correct sample collection, sample storage and sample preparation is of utmost importance to obtain reliable and useful biological information. Many metabolites are highly reactive and unwanted reactions should be avoided. This issue is further addressed in Chapter 2. The main analytical techniques used for the analysis of the metabolome are nuclear magnetic resonance spectroscopy (NMR) and hyphenated techniques, such as gas chromatography (GC) and liquid chromatography (LC) coupled to mass spectrometry (MS). Obviously, other techniques are also possible, e.g. capillary electrophoresis coupled to MS, LC coupled to electrochemical detection, Fourier transform infrared spectroscopy and direct infusion
mass spectrometry. However, none of the individual analytical methods allow us to analyze the full range of metabolites present in cells.

![Figure 3 Schematic of the two-dimensional gas chromatography mass spectrometry system used; PTV: programmed-temperature vaporizer injector, M: interface consisting of a dual-stage four-jet modulator, MS: time-of-flight mass spectrometer equipped with an electron ionization interface.](image)

**Comprehensive gas chromatography mass spectrometry**

Although the general focus in analytical method development lies mostly on LC-MS technology, in metabolomics and related research gas chromatography mass spectrometry (GC-MS) was and still remains a key technology. GC-MS is very suitable for comprehensive analysis due to the high separation efficiencies, repeatable retention times and sensitive and selective mass spectrometric detection. Furthermore problems with ion suppression, a common problem in LC and CE-MS analysis, are virtually non-existing. In addition, the assignment of the identity of peaks detected with GC-MS using electron ionization (formerly known as electron impact ionization) via a database of mass spectra is straightforward, due to the extensive and repeatable fragmentation patterns obtained. However, many metabolites contain polar functional groups and are thermally labile at the temperatures required for their separation or are not volatile at all. Therefore, derivatization of the compounds prior to GC analysis is necessary.

During the last decade a more sophisticated GC technique has become available; comprehensive two-dimensional gas chromatography coupled to mass spectrometry (GC×GC-MS). In Figure 3 a schematic of the GC×GC-MS system used in this research is shown. The principle of GC×GC is based on the coupling of two analytical columns with different selectivities connected through a modulator. The so-called dual-stage cryogenic modulator equipped with four jets (two liquid-nitrogen cooled and two hot gas jets) allows for the consecutive trapping, cryogenic focussing and release of small fractions from the first column effluent in narrow bands onto the second column. In this comprehensive setup the entire sample is separated on both columns and little information from the first separation is lost during the second one. This is achieved by coupling of a first-dimension (1D) capillary column, typically 15 – 30 m long with an internal diameter of 0.25 – 0.32 mm and a film thickness of 0.25 – 1 µm with a short
second-dimension (2D) column, typically 1 – 2 m with an internal diameter of 0.1 – 0.18 mm and a film thickness of 0.1 – 0.18 µm. The first-dimension separation is relatively slow (45 – 120 minutes) compared to the separation on the second-dimension column (1 – 10 seconds). Therefore, the separation on the second-dimension column can be performed simultaneously with the first dimension separation, without losing resolution in the first dimension.\(^\text{10}\) The resulting GC×GC chromatogram consists of a large series of consecutive second-dimension separations. For easier interpretation, these are usually stacked side-by-side to form a two-dimensional chromatogram with the first dimension retention time on the x-axis and the second dimension retention time on the y-axis. Colour gradients are then used to indicate the abundance of individual peaks (Figure 4).

Two-dimensional gas chromatography offers several advantages over one-dimensional gas chromatography, i.e. higher separation efficiency, a larger dynamic range and lower detection limits. However, quantification of all metabolites in metabolomics samples using GC×GC-MS is still a major challenge, due to the many steps required to obtain quantitative data in GC×GC-MS (deconvolution, peak picking, peak merging, and integration) and the complexity of typical metabolomics samples. As an example, thousands of metabolites were detected in a single GC×GC-MS chromatogram of a derivatized plant extract (Figure 5).
General introduction and scope

Chapter 1

Figure 4 Generation and visualization of a GC×GC chromatogram (with permission from Jens Dallüge\textsuperscript{10}).

Figure 5 Two dimensional colour plot of a full scan GC×GC-MS chromatogram of a plant extract (traditional Chinese medicine) after oximation and subsequent silylation.
Scope
Reliable, quantitative analytical methods are prerequisites for correct biological interpretation of metabolomics data. In this thesis the development of quantitative GC-MS based methods for metabolomics analysis is described. In Chapter 2 the challenges in comprehensive (non-targeted) GC-based metabolomics analysis are discussed and recommendations are given on how to validate comprehensive methods from sample extraction up to data preprocessing and how to perform quality control during metabolic studies. Furthermore, an overview is given of method-validation and data-preprocessing methods used in published literature. In Chapter 3 the development and validation is described of a gas chromatography-mass spectrometry method, consisting of an oximation and silylation derivatization reaction and subsequent analysis with GC-MS. This method allows the analysis of a large range of small medium-polar to polar metabolites in cells. In Chapter 4 the differences in method performance for different metabolites from different compound classes were investigated and strategies to improve the precision of the analysis of derivatized metabolites are proposed and demonstrated. In Chapter 5 the development and system performance of a two-dimensional GC×GC-MS method is described. By using a wider bore thicker film 2D column, the most important drawbacks of the conventionally used narrow bore thin film 2D columns were overcome, e.g. the limited mass loadability and limited inertness towards the metabolites of interest. In Chapter 6 the optimized GC×GC-MS method was applied to a set of mouse-liver samples and the feasibility of using commercially available software for automated non-target processing of GC×GC-MS data was assessed. In Chapter 7 the optimization and method performance of an in-liner derivatization method coupled on-line to a GC-MS system is described. The method was suitable for metabolic profiling in ultra-small sample volumes of one microlitre down to aliquots of 10 nL of serum sample. With the optimized method the first comprehensive GC-MS metabolite profiles could be obtained from an aliquot of intracellular content of a single Xenopus laevis cell. In Chapter 8 general conclusions are drawn and future perspectives are discussed.
Reference List

6. Forster, J.; Famili, I.; Fu, P.; Palsson, B. +.; Nielsen, J. Genome Res. 2003, 13 (2), 244-253.