

**Systematic metabolite annotation and identification  
in complex biological extracts**

**Combining robust mass spectrometry fragmentation  
and nuclear magnetic resonance spectroscopy**

**Justin J.J. van der Hooft**

## **Thesis committee**

### **Promoters**

Prof. dr. R.J. Bino  
Professor of Metabolomics of Plants  
Wageningen University  
Prof. dr. S.C. de Vries  
Professor of Biochemistry  
Wageningen University

### **Co-promoters**

Dr. ir. J.J.M. Vervoort  
Associate professor, Laboratory of Biochemistry, Wageningen University  
Dr. R.C.H. de Vos  
Senior scientist, Business Unit Bioscience, Plant Research International,  
Wageningen University and Research Centre

### **Other members**

Prof. dr. H.J. Bouwmeester, Wageningen University  
Prof. dr. T. Hankemeier, Leiden University  
Prof. dr. D. Rolin, University of Bordeaux, France  
Prof. dr. M.W.F. Nielen, Wageningen University

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

**Systematic metabolite annotation and identification  
in complex biological extracts**

**Combining robust mass spectrometry fragmentation  
and nuclear magnetic resonance spectroscopy**

**Justin J.J. van der Hooft**

**Thesis**

submitted in fulfillment of the requirements for the degree of doctor  
at Wageningen University  
by the authority of the Rector Magnificus  
Prof. dr. M.J. Kropff,  
in the presence of the  
Thesis Committee appointed by the Academic Board  
to be defended in public  
on Thursday 4 October 2012  
at 11 a.m. in the Aula.

Justin J.J. van der Hooft

Systematic metabolite annotation and identification in complex biological extracts -  
Combining robust mass spectrometry fragmentation and nuclear magnetic  
resonance spectroscopy,  
256 pages.

Thesis, Wageningen University, Wageningen, NL (2012)

With references, with summaries in Dutch and English.

ISBN 978-94-6173-234-7



## *Table of Contents*

Preface	9
Chapter 1: General Introduction	11
Chapter 2: Polyphenol Identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation	39
Supporting information Chapter 2	63
Chapter 3: Spectral trees as a robust annotation tool in LC-MS based metabolomics	79
Supporting Information Chapter 3	107
Chapter 4: A strategy for fast structural elucidation of metabolites in small volume plant extracts using automated MS-guided LC-MS-SPE-NMR	121
Chapter 5: Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts	139
Chapter 6: Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake	165
Supporting information Chapter 6	197
Chapter 7: General discussion	207
Chapter 8A: Summary	229
Hoofdstuk 8B: Samenvatting (NL)	235
List of abbreviations used	241
Acknowledgements	245
Curriculum Vitae	249
Publication list	250
List of co-author affiliations	252
Overview of the completed training activities	255



*To the beautiful world  
and all those people that dare to ask  
how it works*



## *Preface*

The understanding of the complex physiology of organisms asks for dedicated tools to annotate and identify the molecular cell components that comprise of DNA, RNA, protein, and small metabolites. Protocols for detection and subsequent characterization and identification of DNA, RNA, and proteins are quite well established nowadays, while the tools for metabolite annotation and identification are still in development. Metabolomics, which is the scientific discipline that encompasses the tools that aim to fully describe the set of small metabolites, still searches for strategies for the fast identification of metabolites out of a minute amount of sample. My PhD project aims to establish novel methods and protocols for the unambiguous identification of yet unknown metabolites by combining high resolution mass spectrometry (LC-MS<sup>n</sup>) and nuclear magnetic resonance (NMR) spectroscopy methods, starting with plant derived semi-polar compounds and focusing on phenolic compounds. The first chapter serves as introduction into the techniques and samples used throughout this thesis. The three subsequent chapters describe the development and implementation of the two analytical platforms, based on the analytical techniques explained in the introduction of this thesis. These platforms were then applied to plant-derived compounds and human metabolized compounds as is reported in the experimental chapters 5 and 6. The discussion of this thesis reviews the current developments in metabolite identification in view of the results published in this thesis.

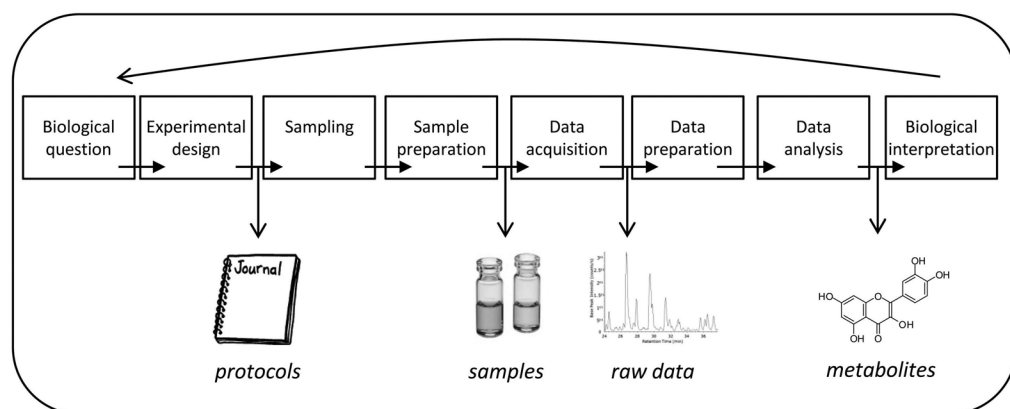


## *Chapter 1: General Introduction*

To better understand physiological processes in plants and humans, a more detailed understanding of the biochemical processes within organisms, organs, tissues, and cells is highly desired. The exact chemical entity of small molecules (i.e. metabolites) present in organs, tissues and cells is required in order to understand the growth and metabolism of all organisms. However, the chemical variety of molecules present in biological systems is enormous. Metabolomics is often referred to as the discipline that aims to get a comprehensive picture of all the small biomolecules, i.e. metabolites, present in each organism.<sup>1</sup> To identify the exact chemical structure of the large number of metabolites present in biological extracts is one of the major obstacles in metabolomics research. Actually, no single analytical platform exists that can measure and identify all existing metabolites. To solve this bottleneck in metabolite identification, improved analytical protocols and data analysis strategies are needed.

The overall aim of this thesis is to improve the metabolite identification process by developing and combining mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR)-based methods in a way that robust spectral data is generated that can be used as metabolite-specific fingerprints and that provide structural information of the detected compounds at the same time. Subsequently, it is intended that the systematic (and preferably semi-automated or even fully automated) analysis of these spectral data can lead to the complete structural elucidation of the metabolites involved. In the chapters two to four, the developed analytical platforms are discussed, and in chapter five and six these platforms are used to solve metabolite structures in biological extracts from plant and human origin.

This PhD project is embedded within one of the core themes of the Netherlands Metabolomics Centre (NMC; [www.metabolomicscentre.nl](http://www.metabolomicscentre.nl)), a centre which focuses on the development of metabolomics based technologies and instrumentation to address the current and future challenges in biology, biotechnology and biomedical research in order to improve personalized health



**Figure 1:** Metabolomics pipeline as used within the Netherlands Metabolomics Centre. The end of the pipeline is connected to the start, since the biological interpretation of the data usually raises new questions about the biological function of identified metabolites, how they fit into existing metabolic networks, or factors that influence the presence of the discovered metabolites.

and quality of life. The NMC consortium includes both core projects, which are research-driven and/or method-driven, and associate projects, which are in tight collaboration with industries aiming for tailor-made analytical solutions. Thereby, a platform for metabolomics studies is created that provides tools for an improved metabolomics pipeline (Figure 1). The role of metabolite identification in this pipeline is to aid in the biological interpretation of metabolomics data sets and thereby creating new hypothesis for more biological questions. This PhD project focuses on a more robust, sensitive, and automated identification of metabolites from plant and human origin using multistage MS fragmentation ( $MS^n$ ) and NMR spectra. Both types of spectral data will then serve as input for other projects active in the NMC frame work.

The chemical analysis of plant derived food products and their metabolites in the human body represents an increasingly important field in biology, as it forms the basis to gain deeper insight into the physiological responses to food intake. After all, we all need food as our daily energy intake. The consumption of phenolic compounds present in food and beverages is of particular interest due to the potential health benefits in relation to cardiovascular diseases that are associated with intake of polyphenols, a compound class widely distributed in the plant kingdom (see also section on phenolic compounds).<sup>2</sup> The bioavailability of



phenolic compounds has been studied by many groups over the latest years.<sup>3</sup>  
<sup>4</sup> Polyphenols are partly absorbed in the upper part (small intestine) of the gastro-intestinal tract, but it is generally accepted that the majority of phenolic compounds reach the colon where microbes can break down the phenol cores before uptake into the blood and subsequent conjugation in the liver. Therefore, the scientific community is increasingly aware that not only conjugated phenolic molecules, but also conjugated microbial-derived metabolites can contribute to the health effects associated with polyphenol intake. Some of these metabolites are still excreted into the urine at more than 24 hours after a single intake of polyphenols, which indicates the large extent of metabolism that can occur before excretion into the urine. The identity of the chemical entities present in the human body serves as a good start for detailed studies on bioavailability of diet compounds and personal response studies.

Current gaps in metabolite identification are often observed in untargeted large scale metabolomics studies, in which frequently ‘unknown metabolite features’ are found to be discriminating for a certain trait or stage of disease. In these studies, thousands of metabolites are usually detected by MS-based platforms; however, most of these metabolite features can, as yet, usually not be annotated to existing chemical entities. In order to facilitate the annotation and identification process of metabolites present in food and urine extracts, this PhD project focuses on improving sample preparation, spectral data acquisition and data analysis. Especially secondary metabolites of plant origin and conjugated endogenous and exogenous metabolites in urine extracts are difficult to identify, due to a lack of reference compounds and due to the presence of many closely related compounds including isomers, i.e., compounds with the same elemental formula but different chemical structures. As proof of principle for the applicability of MS and NMR based identification platforms, the annotation and identification of phenolic compounds and conjugated derivatives thereof will be used as they represent a diverse compound class yet sharing similar core features. Methanol-water extracts of *Lycopersicum esculentum* (tomato), *Arabidopsis thaliana* and *Camellia sinensis* (tea) were selected, as these are well-studied but as yet far from

completely understood systems. These plant extracts contain a wide array of phenolics, typically glycosylated with sugar moieties like glucose and rhamnose. Also, extracts of human urine, collected after tea intake, were used to annotate and identify phenol-derived and endogenous metabolites, both conjugated by the human host. In fact, metabolite identification in these complex extracts involves dealing with a difficult background matrix and, in case of human body fluids, also low abundant target analytes, which results in considerable time and effort needed by data experts to perform metabolite identifications.

This introduction will discuss details of metabolomics strategies and the general metabolomics pipeline, like sample preparation techniques, principal aspects of data acquisition by mass spectrometry and nuclear magnetic resonance spectroscopy, and computational tools to analyze spectral data, thereby focusing on metabolite identification. Finally, an overview of the thesis chapters is provided in which my contributions are mentioned.

*Metabolomics: atoms as building blocks.*

Metabolomics is the youngest member of the *~omics* family that traditionally consists of genomics, transcriptomics, and proteomics. Considerable progress has been made in the maturation of analytical tools in the different *~omics* fields, as, for example, resulted in integrated workflows for protein identification and quantification in proteomics.<sup>5</sup> In metabolomics, the basic building blocks are atoms, and in fact, most biological molecules consist of a few types of atoms with carbon and hydrogen being the most abundant atoms. The real challenge of metabolite identification starts with the endless amount of possible configurations when building molecules out of carbon, hydrogen, oxygen, nitrogen, phosphor, and sulphur atoms. On the basis of all the existing configurations and the vast amount of data sets available, it is estimated that the amount of metabolites present in plants is 1.000.000 or even more.<sup>6</sup> This huge amount of metabolites results in an enormous diversity in compounds from being small in molecular size (100 Dalton (Da) or lower, like ethanol) up to larger complex molecules (1000 – 1200 Da, like heavily glycosylated polyphenols), and being either polar (sugars),

semi-polar (polyphenols and their glycosides) or apolar (lipids). Theoretically, up to  $50 \cdot 10^9$  different metabolites exist of 1200 Da or lower, of which currently a very small percentage is present in compound databases.<sup>7</sup>

*Phenolic compounds: the colorful compound class of flavonoids.*

Phenolic metabolites are molecules that contain at least one hydroxybenzene (i.e. phenol) moiety as part of their structure, like, for instance, the flavonoids. These molecules are widely distributed throughout the plant kingdom, as is illustrated by the presence of more than 6000 different flavonoids.<sup>8</sup> Flavonoids have different biological functions in plants, including protection against biotic and abiotic stresses, pigmentation, and signaling.<sup>9</sup> Flavonoids are so-called polyphenols as they consist of a core structure containing two hydroxylated benzene rings connected by a heteroatom ring. Depending on the location of hydroxyl groups and the type of linkage in the heteroatom ring, different flavonoid subclasses are defined, for example flavones and flavonols. The chemical complexity of flavonoids is caused by the many possible hydroxylation sites, as well as the different glycosylation, acylation or methylation patterns that can be present on one or more hydroxyl groups, leading to the wide array of flavonoids.

*Phenolics in food: determining the metabolic fate of phenols upon human intake.*

Dietary intake of (poly)phenolics present in food and beverages has been associated with health benefits for humans, for example in relation to cardiovascular diseases.<sup>2</sup> This makes it of particular interest to characterize individual species of this class of compounds present in crude extracts from plants. For example, leaves of the tea plant *Camellia sinensis* have been used for ages to prepare different kind of tea drinks. The tea plant contains numerous polyphenolic compounds and derivatives, of which the flavan-3-ols catechin and epicatechin belong to the most abundant metabolites.<sup>10</sup> There is an increasing scientific interest in the bioavailability of phenols from tea and other sources and their fate after dietary intake.<sup>3,4</sup> In the human body, phenol-derived metabolites

are mostly present as conjugates with methyl, sulphate, and/or glucuronide moieties. The biotransformation of food-derived molecules like phenolics is catalyzed by so-called phase I and phase II enzymes that are, for instance, present in cells in the small intestine and especially in the liver. The exact structures of these metabolized phenolics is difficult to elucidate, due to their relative low abundance, i.e., concentrations of about 2 nM in plasma and 2  $\mu$ M in urine. The analytical platforms to be developed within this PhD project will be used for identification and annotation of both (poly)phenolics and their metabolized products.

#### *Sample preparation techniques in metabolomics.*

Sample preparation is a major facet of metabolomics studies, as the quality of the generated data very much depends on the quality of the analyzed material. The area of sample preparation techniques in metabolomics has significantly developed over the last decade.<sup>11</sup> A general and robust sample preparation, i.e. extraction protocol, is desired for reproducible measurements, which enables comparison of data sets from larger series of samples. For metabolite identification, the concentration and purification of compounds from complex matrices results in higher-quality and purer metabolite spectral data, which facilitates the extraction of relevant structural information from the generated data. Different strategies are employed, from liquid-liquid extractions to Solid Phase Extraction (SPE) protocols, which works on the principle of retaining the compounds of interest on a solid material (i.e. sorbent) after which the matrix is washed away. A robust and fast methanol-water extraction of plant material resulting in crude semi-polar extracts suitable for LC-MS measurements was described by De Vos et al.<sup>12</sup> Phenolic molecules were shown to be reproducibly and well extracted, thus the protocols described in that paper will be used as starting point for identifying plant phenolic metabolites within this thesis.

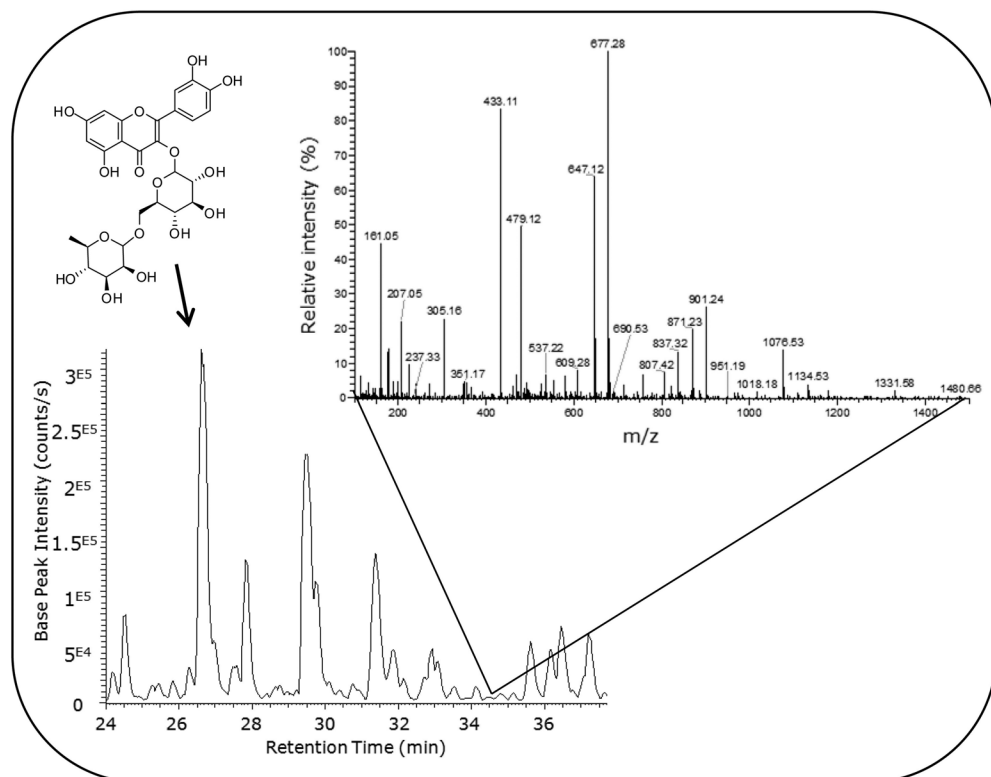
*Separating metabolites: chromatography in metabolomics.*

In case of complex samples, separation of metabolites, in particular structurally related compounds or isomers, enables their individual detection and provides additional structural information, thereby greatly facilitating metabolite profiling. Separation of metabolites can be achieved in different ways and is dependent on the physical and chemical characteristics of the molecules, i.e. volatile/non-volatile, polar/apolar, etc.<sup>13</sup> In case of volatile organic compounds, like monoterpenes, short-chain aldehydes and alcohols, or polar compounds that can be made volatile by derivatization, like sugars and organic acids, gas chromatography (GC) coupled to MS represents a well-established analytical tool, making use of the stable GC and the robust high-energy electron impact spectra generated for database searching, e.g. in the NIST library (see also section on fragmentation in MS).<sup>14</sup> High pressure/performance liquid chromatography (HPLC) is currently the most used technique for separating semi-polar compounds (non-volatiles), such as the (poly)phenolic compounds present in food derived extracts, as these compounds cannot be separated with GC. The actual separation with LC is performed using a gradient of solvents that cause metabolites to retain for different times on solid column material. Different types of sorbents (column material) exist, for example hydrophilic interaction liquid chromatography (HILIC) sorbent that is suitable for separation of polar metabolites (also referred to as normal phase) and octadecyl (C18) material that is suitable for semi-polar to apolar compounds (also referred to as reversed phase). The quite recently introduced ultra high performance/pressure liquid chromatography (UPLC) offers advantages over HPLC such as shorter run times and sharper LC peaks.<sup>15-17</sup> The diameter of the column used for chromatographic separation determines the amount of sample that can be injected into the HPLC system. Columns with a diameter of 4.6 mm or less are usually called analytical columns, while columns with a higher diameter are called semi-preparative or preparative columns. Which type of column to use depends on the nature of the experiment, that is, for metabolic profiling the smaller sample volumes and well defined peak shapes of analytical columns are preferred over the usually broader and less well

defined peak shapes of preparative columns. This peak broadening in preparative columns is due to the increased column volume the compounds experience, which gives compounds more time to diffuse before elution. For metabolite identification purposes, usually higher volumes are preferred to collect as much as possible compound or compounds of interest that need to be identified. The semi-analytical 4.6 mm columns represent a good compromise between injection volumes and LC peak shapes.

*Mass spectrometry in metabolomics.*

Mass spectrometry (MS) has become the principal analytical technique in metabolomics. Compared to a photodiode array detector (measuring ultra violet (UV)-visible light absorption) and a fluorescence detector, mass spectrometry can analyze both chromophores and non-chromophores or fluorophores and non-fluorophores. Molecules that enter the source of a mass spectrometer get in the gas phase and then get charged at high voltage, which enables their detection with a mass analyzer. The mass analyzer then measures mass (to charge) values to which elemental formulas can be assigned using the mass of individual atoms. In MS only charged molecules, i.e. ions, can be detected as metabolites unable to ionize will not reach the detector. In GC-MS, hard ionization techniques like electron impact (EI) are suitable for robust fragmentation and detection of ions<sup>15</sup>. However, in combination with LC, electron impact MS does not efficiently work due to the presence of the elution solvents upon during the ionization. Therefore, soft ionization techniques were introduced as a suitable alternative approach to ionize LC-separated molecules.<sup>16</sup> The detection of molecules in soft-ionization MS, like electrospray ionization (ESI), is dependent on the capacity of the molecule to ionize while being part of a complex mixture. In case of complex crude extracts, MS can detect many mass peaks within a small abundant LC-MS peak (Figure 2). Apart from the chemical properties of a molecule itself, the eluent flow and composition, sample matrix and ionization source all influence the molecules' ionization.<sup>18</sup> Therefore, so called ion suppression and matrix effects can hamper robust detection of molecules that ionize less well. This is



**Figure 2:** Part of LC-MS chromatogram of an extract from *Lycopersicum esculentum* (tomato) where flavonoid glycosides elute, with the base peak intensity on the y-axis and the retention time on the x-axis. Several abundant base peaks are visible, i.e., the peak at 26.8 min corresponds to rutin (= quercetin-3-O-rhamnosyl-1,6-O-glucoside), and the displayed m/z chromatogram (relative intensity vs. m/z) shows the complexity of crude extracts as even in minor peaks multiple metabolites elute displaying their parent and in-source fragment ions.

particularly hampering robust MS-based quantification on the basis of LC-MS peak abundances or areas, as co-eluting signals can influence the MS signal abundance and peak shape of a metabolite. Nevertheless, one of the advantages of soft ionization is that it allows for sensitive detection of the pseudo-molecular ion, which is useful for metabolite identification.

Different types of mass analyzers exist, each with their own advantages and disadvantages,<sup>18-20</sup> resulting from the differences in accuracy, speed, resolution, and dynamic range. Typically, the mass accuracy of measured ions is expressed in parts per million (ppm) of the difference between observed and theoretical mass. Mass accuracy is an important parameter as it makes it easier to assign

elemental formulas (EFs) to the observed mass features. The higher the mass accuracy, the lower is the amount of possible elemental formulas for a given  $m/z$  value. Additional information like natural isotope abundance is very valuable to further minimize the number of candidate EFs,<sup>21</sup> as, for instance, a mass spectrometer with a 3 ppm mass error and 2% error in isotope abundance patterns outperforms a very accurate mass spectrometer (0.1 ppm) that does not include isotope abundance patterns in the analysis. Time of Flight (TOF) type mass analyzers offer fast scanning times and a reasonably dynamic range, but their mass accuracy is modest (usually about 5 ppm). Ion trap analyzers are fast scanning detectors, but have the substantial drawback of obtaining nominal mass accuracy. During fragmentation experiments in the mass spectrometer, however, an advantage of using an Ion trap compared to a TOF analyzer is its possibility to trap not only parent ions but also its fragment ions, and subsequently perform multi-stage mass fragmentation ( $MS^n$ ), thereby gaining insight into the precursor – product fragment ion relationships and thus having the ability to discriminate metabolites based on differential fragmentation pathways.

In terms of a high mass accuracy ( $> 60,000$ ) and a wide dynamic range (i.e., up to 20000 Da and higher), there are currently two mass analyzers available that are based on Fourier Transformation of circulating ions, namely the ion cyclotron resonance (FT-ICR-MS) and the Orbitrap type (FT-Orbitrap-MS, or Orbitrap-FTMS). A FT-ICR-MS can reach the highest mass resolving power so far reported for any mass spectrometer ( $>1,000,000$ ) with a mass accuracy generally within 1 ppm.<sup>22</sup> The recently developed Orbitrap-FTMS has, as compared to the FT-ICR-MS, a more modest performance (maximum resolving power  $> 200,000$  and within 2 ppm of mass accuracy with internal standard), but has lower instrument costs, and the Orbitrap instrument is relatively faster due to shorter scan times. The latter is very advantageous when hyphenated to separation techniques, such as LC, and also during fragmentation experiments.<sup>18, 23, 24</sup>

In short, LC-MS based platforms can yield a large breath of information on biological extracts, usually detecting in each sample thousands of so-called mass features comprising of parent ions, in-source fragments, and adducts from



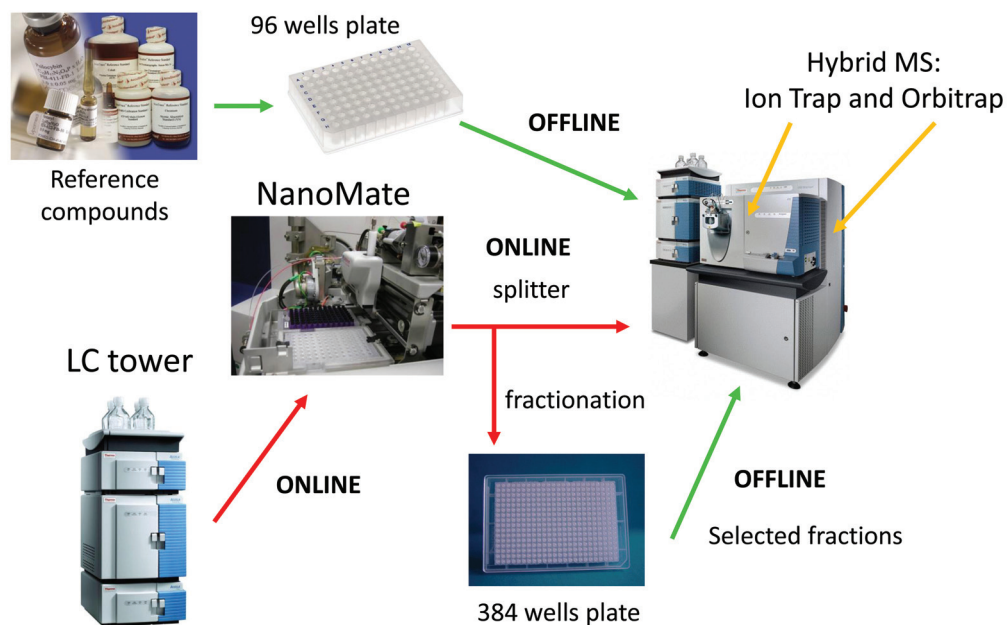
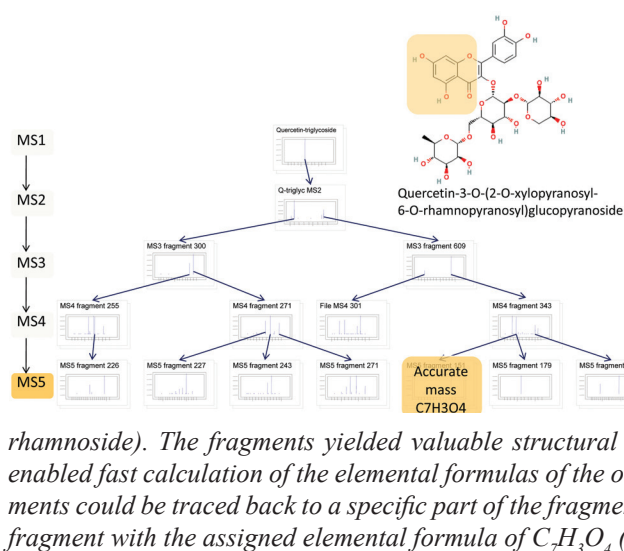
several hundreds of metabolites. The use of accurate mass based platforms not only increases the separation power of isobaric molecules, but also facilitates fast elemental formula (EF) assignment to detected metabolite features, which is one of the key steps toward identification of a metabolite.<sup>7</sup> For instance, an analytical HPLC-QTOFMS based platform was developed to characterize *Lycopersicum esculentum* (tomato) metabolites and store metabolite features in a database.<sup>25</sup> Additional structural information can be obtained from MS experiments by performing fragmentation experiments in the mass spectrometer, as discussed in the next section.

*The use of fragmentation in mass spectrometry.*

Fragmentation of metabolites within the mass spectrometer can yield additional structural information by detection of compound class specific fragments or losses. Therefore, the use of LC-MS/MS and tandem MS approaches has been advocated to obtain further details on metabolite structures.<sup>19,20</sup> Metabolites can be fragmented by adding collision energy to trapped molecular ions, for instance by using inert gases like argon. A regular type of fragmentation is collision induced dissociation (CID) that uses inert gas molecules to increase internal energies of parent ions.<sup>19</sup>

A direct comparison of fragmentation spectra obtained from collision-induced dissociated metabolites in LC-MS can be challenging, as the relative low energy used for fragmentation, as compared to EI-MS, generally results in variable fragmentation spectra.<sup>19</sup> Robust and reproducible fragmentation patterns are advantageous to compare metabolite fragmentation patterns and thus for database storage, as can be seen in GC-MS that successfully uses reference databases for metabolite identification.<sup>14</sup>

Multistage mass spectrometry (MS<sup>n</sup>) fragmentation is based on the trapping of fragment ions and subsequent fragmentation of these fragment ions in an Ion trap, resulting in higher level MS fragmentation spectra. If done in a systematic way, so-called spectral trees can be formed,<sup>26</sup> as is illustrated by the fragmentation of a complex plant polyphenol glycoside in Figure 3. Different fragmentation paths



become visible, because the precursor – daughter ion relationship is recorded. This is in contrast to LC-MS/MS (e.g. TOF or Ion trap) at higher collision energies where the fragments are detected but their relations are not directly measured. Some fragments can be traced back to a specific part of the fragmented molecule (Figure 3), thereby providing structural information.

Within this project, a chip-based nano-electrospray ionization source (Advion Nanomate Triversa) will be used in combination with two consecutive mass analyzers (Figure 4), namely an Ion trap MS, providing reproducible fragmentation, and an Orbitrap FTMS providing accurate mass read-out and thus enabling rapid assignment of elemental formulas to the molecular ions and all fragment ions derived thereof. The fragmentation approach will be tested with reference compounds and in combination with LC of crude extracts (Figure 4). This set up enables detailed offline MS<sup>n</sup> analysis in either negative or positive ionization modes, or both, if additional structural information is needed to discriminate and identify the target compound. During this PhD project, MS<sup>n</sup> fragmentation approaches with the NanoMate robot will be developed using the identification and annotation of the compound class of (poly)phenols as an example.

Thus, both full scan MS and MS-fragmentation are very versatile tools in metabolite detection, selection, annotation and structural elucidation. Nevertheless, for complete identification of unknown compounds, the structural information from nuclear magnetic resonance (NMR) spectroscopy is often required.

*Solving the chemical puzzle with Nuclear Magnetic Resonance spectroscopy (NMR).*

NMR is an analytical tool that measures energy differences in spin containing particles that are subject to a magnetic field. Because of this principle, NMR signals are a direct reflection of the amount of spin particles present and hence NMR is a quantitative method. <sup>1</sup>H and <sup>13</sup>C are the most common spin particles used in NMR. NMR measurements provide structural information about the

chemical environment (shifts, normally represented in ppm values) of spin-containing particles and the number of neighboring spin particles (Figure 5). Further details that can be observed in a NMR spectrum are different coupling constants ( $J$ , normally presented in Hz) that are characteristic for certain proton–proton interactions, thereby providing structural information. A magnetic field strength of 14.1 Tesla, corresponding to 600 MHz ( $^1\text{H}$ ) spectrometers, usually provides enough resolution to correctly assign coupling constants of protons present in small molecules. The structural information obtained by NMR can define, for instance, the exact attachment side of a sugar moiety on the molecular core. Extensive NMR measurements show connections between groups of connected protons, or between protons and carbons, present in a metabolite. In this section, NMR spectroscopy and NMR data acquisition for metabolomics experiments will be introduced.

One dimensional (1D)- $^1\text{H}$ -NMR is the most sensitive NMR application,<sup>18</sup> due to the high natural abundance of hydrogen atoms in molecules and the relative high sensitivity of NMR to the  $^1\text{H}$  isotope as compared to other NMR detectable isotopes like  $^{13}\text{C}$ . NMR-based compound identification is most straightforward on the basis of direct comparison of proton signals of 1D- $^1\text{H}$ -NMR experimental spectra with those of reference spectra. However, in the process of de novo metabolite identification, more structural information than can be derived from 1D- $^1\text{H}$ -NMR spectra is usually needed to find and prove the correct chemical structure. For example, to determine how (groups of) protons are connected to each other within the molecule, two dimensional (2D)- $^1\text{H}$  and 2D heteronuclear experiments, for instance measuring proton – carbon connections, are useful tools to collect more structural pieces of information to complete the molecular puzzle. However these type of measurements need additional experimental time and expertise to interpret the data.<sup>18</sup>

A NMR experiment is performed by applying pulse programs that define the possible energy transfers and mixing times and thus the type of NMR experiment that is conducted, for instance, regular 1D- $^1\text{H}$ -NMR or 2D connections through bond or through space. By applying specific pulses, the type of detectable

signals can be regulated. For example, by saturating the protons attached to non-deuterated water molecules (water suppression), they become much less visible in the NMR spectrum. This is especially helpful in the study of metabolites in biological extracts, because it is very difficult to get rid of all water in such extracts before redissolving in deuterated solvents suitable for NMR (Figure 5). Using water suppression, a larger receiver gain of the NMR spectrometer can be used without the risk of overload, thereby revealing more of the relevant NMR signals.

The sensitivity of NMR spectroscopy is still relatively low as compared to MS: while MS can detect metabolites at picomol to femtomol ( $10^{-12}$  -  $10^{-15}$  mol) amounts, NMR can detect down to low nanomol ( $10^{-9}$  mol) amounts. Over the latest years remarkable progress has been made in decreasing the detection limits for NMR. These technological advancements include the introduction of probes that are compatible with smaller volume NMR tubes and a cryogenically cooled probe head (sensitivity gain of a factor 4), which is especially useful in mass limited samples.<sup>27</sup> With the increased sensitivity of NMR detection systems, background signals arising from non-deuterated solvents and contaminants like co-eluting compounds also increase using a more sensitive NMR approach. In case of large deuterated solvent peaks, this can be partly overcome by abovementioned (water) suppression pulses; however, structural information of protons resonating in the neighborhood of the suppressed region might get lost. Thus, the purity of the sample is of utmost importance if only trace amounts are available for structural elucidation. In practice, with 30  $\mu\text{g}$  (300  $\mu\text{M}$  in 3 mm NMR tube for MW of 500) of (>80%) pure compound in the probe, a full set of 1D and 2D-NMR experiments can be performed within a reasonable amount of time, i.e., a heteronuclear 2D-experiment detecting  $^1\text{H}$ - $^{13}\text{C}$  bonds over a longer range (HMBC), the most time-consuming 2D-NMR experiment, will take about 24 hours. At lower amounts, i.e. 5-10  $\mu\text{g}$  (50-100  $\mu\text{M}$ ), homonuclear 2D experiments (i.e., measuring  $^1\text{H}$ - $^1\text{H}$  bonds over three bonds [COSY] or within a group of protons [TOCSY], being less time-consuming) are still feasible within a few hours.<sup>18</sup> The experimental time to obtain NMR spectra with sufficient signal-

**Figure 5 (on next page):** Proton NMR spectrum from a sample in a flow probe obtained by LC-SPE-NMR. The spectrum is divided into the aromatic region (left) where protons attached to phenolic rings are visible in the spectrum, and the sugar region where protons attached to the sugar moiety are visible. On top left, two double doublets from the sugar region are shown with two marked coupling constants  $J$  (normally indicated in Hz). A Bruker 600 MHz spectrometer, used to obtain these spectra, is displayed as well.

to-noise will increase with lower amounts of analyte. Hence, often compromises have to be made between available measurement time, compound purification, and the amount of structural information needed.

*MS and NMR as complementary analytical techniques.*

Both MS and NMR have their own advantages and disadvantages in the use for metabolite detection and identification.<sup>18</sup> A regular LC-MS run usually takes about one hour, depending upon the column used and the gradient applied. Due to the LC-separation of the metabolites in a crude extract and the selectivity of MS, a substantial amount of structural information can be retrieved. However, MS/MS, or even MS<sup>n</sup> spectra, once obtained, still may not be sufficient to completely solve the metabolite structure. A regular 1D-<sup>1</sup>H-NMR analysis takes about 15 minutes. More extensive 2D-NMR experiments can take 2 hours up to 24 hours, depending on the amount of analyte present in the probe and the chosen NMR experiment. In case of pure compounds, the generated NMR data sets contain valuable information needed for fulfillment of the metabolite structure. A mixture of compounds, however, may result in difficulties to interpret the NMR data.

Apart from the higher sensitivity of MS, advantages of MS-based spectral data over NMR-based spectral data include i) a high resolving power, providing the possibility to detect co-eluting compounds that have different parent masses, and ii) the possibility to fragment metabolites to obtain substructure information. In the NMR probe, compounds are difficult to separate during analysis, especially if they are present at similar abundances. Likewise, NMR offers some unique structural information as compared to MS. For example, the stereochemistry of metabolites cannot be determined solely based on MS, while NMR can usually not only discriminate, but also identify (dia-)stereoisomers. However, structural



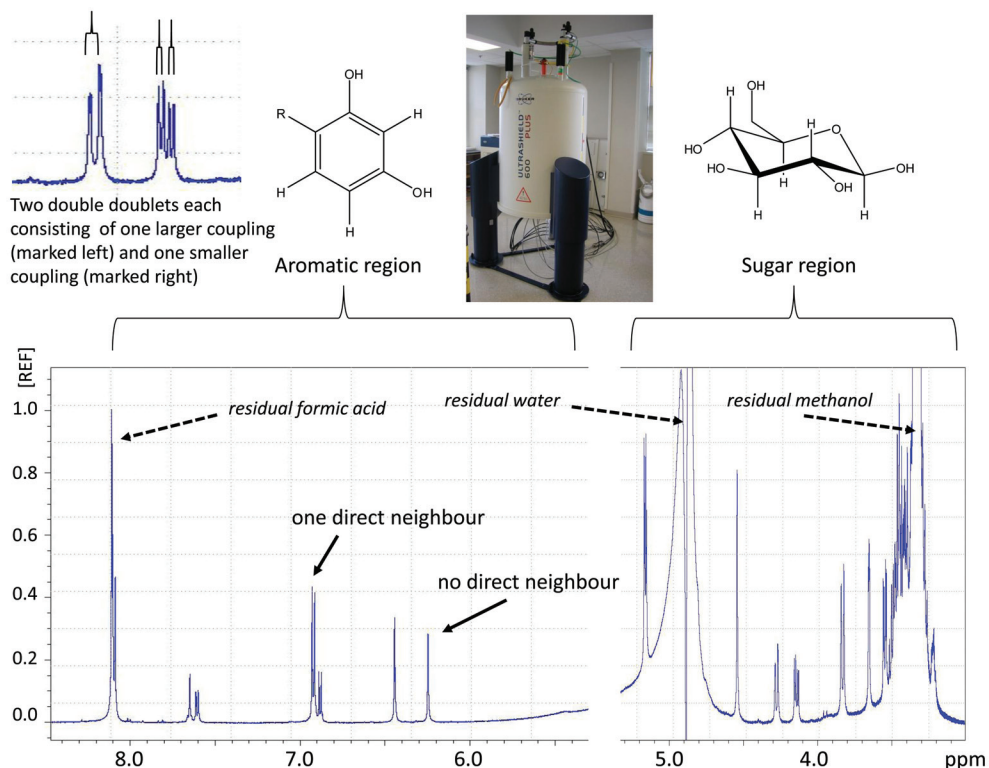


Figure 5: See previous page.

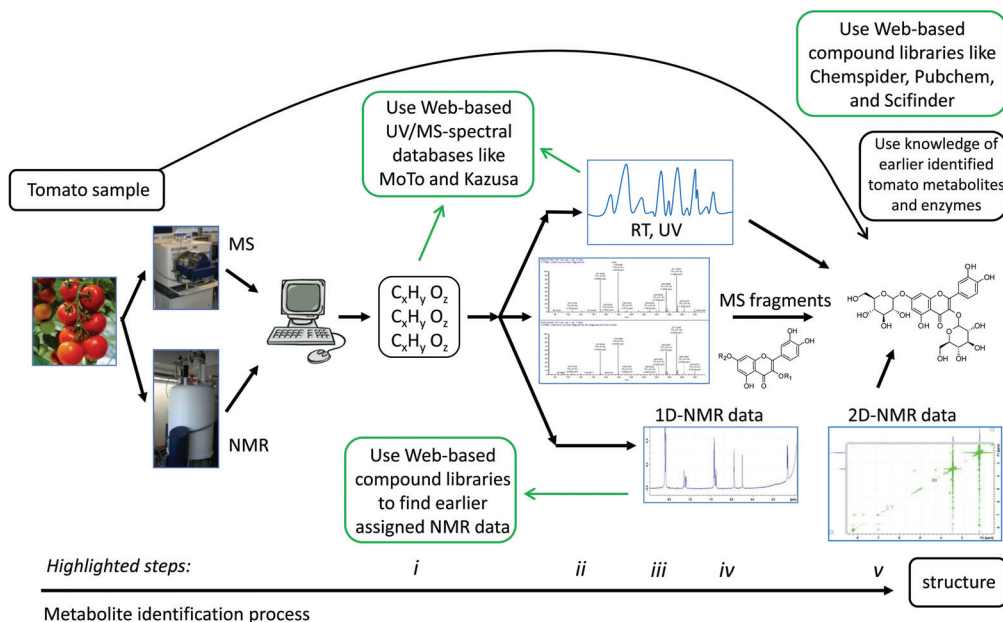


Figure 6: Metabolite identification process using spectral data from LC-UV-MS<sup>n</sup> and NMR.

knowledge from MS is sometimes essential for fast metabolite identification. For example, the functional carboxyl (COOH) and sulphate (SO<sub>3</sub>) groups are not directly visible in <sup>1</sup>H-NMR; however, the presence of these functional groups is usually easily determined with MS. These differences underline that MS and NMR are complementary techniques that can be combined for efficient metabolite identification.

*Bridging the gap between LC and NMR: SPE as connector.*

The so-called hyphenation of LC and NMR (i.e., LC-NMR) is not straightforward, due to the non-deuterated solvents used in regular HPLC which hamper detection in the NMR of the much lower abundant biomolecules.<sup>28,29</sup> The direct coupling of LC and NMR has been made by, for instance, stop-flow experiments using deuterated solvents.<sup>29</sup> The introduction of SPE-NMR enabled an efficient connection between HPLC and NMR by post-column trapping compounds of interest on small SPE cartridges, based on their Ultra Violet (UV) absorption characteristics measured by a Diode Array Detector (DAD). The inclusion of the Prospekt SPE unit (Spark Holland) as part of the hyphenated system enables the removal of non-deuterated solvents by drying the filled SPE cartridges and subsequent elution with deuterated solvents that are suitable for NMR.<sup>28</sup> At the same time, analytes can be (further) concentrated by performing multiple trapping runs, which is useful to collect sufficient analyte for NMR measurements or to decrease the measurement time needed. Structural information obtained from separately run LC-MS and UV guided SPE-NMR hyphenated platforms has been used for complete metabolite identification of, for instance, 22 secondary metabolites from *Kanahia laniflora* and *Harpagophytum procumbens* (Devil's claw)<sup>30,31</sup> and the elucidation of two therapeutic agents using only 15 µgram of analyte isolated from in vitro production with liver microsomes.<sup>32</sup>

The SPE trapping of analytes occurs upon triggering by a signal, for instance the UV absorption at 254 nm where certain flavonoids are visible, which can be set in Hystar software (Bruker). Upon triggering, the eluens flow is redirected from the waste toward a small SPE cartridge. This can be done for multiple peaks in



a chromatogram. In order to improve the recovery of the analytes, a make-up flow is added to the column flow to lower the organic strength of the solution. Over the latest years several SPE sorbents have been developed and tested in the trapping and elution during SPE-NMR measurements, together with other parameters influencing the final amount of analyte present in the NMR (flow) probe.<sup>33</sup>

Not all metabolites show distinct UV absorption and thus cannot be selectively trapped solely on the basis of the UV absorption. MS is often essential for specific metabolite detection and provides structural information useful for metabolite identification. Therefore, within this PhD project, the aim is to further develop the technique of LC-SPE-NMR using a mass spectrometer, instead of a DAD, for triggering SPE-trapping of eluting metabolites.

*Collecting the structural evidence: the Metabolite Identification process.*

In the sections above, analytical tools were described that collect structural information for metabolite identifications. The most straightforward identifications are of those compounds that were run as standard on the same analytical system. However, this is typically only valid for a minor amount of metabolites present in any extract. In-house chemical databases usually contain both fully and partially characterized metabolites and are dedicated toward the specific analytical systems and samples of a laboratory. A nice example is the MoTo database that contains tomato spectral data gathered by a LC-PDA-QTOF-MS system.<sup>25</sup> An advantage of using in-house databases is that generated experimental retention times and mass spectral features can be directly compared to the database values and quick identifications can be made. Those compounds not present in the in-house databases have to be annotated. During metabolite annotation, all available spectral data belonging to one metabolite is combined. On the basis of the gathered structural information, the metabolite identification is started. This usually starts with querying Web-based compound libraries using the annotated spectral data as input. Some libraries contain spectral data of known (reference) compounds, and other databases contain observed mass and/

or NMR features of known and yet unknown metabolites, often from specific biological origin or compound classes.<sup>18, 34</sup>

In literature, different qualifications indicating the level of metabolite identification can be found, like ‘putatively’ or ‘tentatively’ identified, which are used for not completely identified compounds; however, these qualifications were not yet defined, i.e., it remains unclear what these terms exactly mean within the context of metabolite identification. Therefore, the Metabolomics Standards Initiative (MSI) proposed reporting standards for metabolite identifications (MI) in metabolomics studies.<sup>35</sup> The MI levels range from 1, corresponding to a fully elucidated compound, to 4, which is a detected metabolite feature in an extract. A typical metabolite identification process of compounds not present in in-house databases is described here (Figure 6), highlighting five steps and using a flavonoid-diglycoside as example:

- i) The  $m/z$  value of the parent ion is defined and the EF is assigned using LC-MS information (i.e. MSI MI level 4).
- ii) The fragmentation patterns are analyzed and a) it is checked whether the parent ion is true metabolite and no adduct or fragment, and b) additional structural features can be extracted, like the type of aglycone and the amount and type of sugar moieties (i.e. MSI MI level 3). In some cases, the type of linkage between the two sugar moieties can be determined based on MS fragmentation.
- iii) MS-based databases are queried for possible candidates on the basis of EF of parent ions present in studied species, and searches are further refined using the information from retention times (RT), UV absorbance maxima, fragment ions, and other available spectral data. For tomato metabolites, for instance, the MoTo database can be queried (i.e. MSI MI level 2 in case one or only a few similar molecules are found).
- iv) The spectral data from 1D-<sup>1</sup>H-NMR is used to further restrict the amount of candidate structures with the use of core specific NMR patterns and NMR databases (i.e. MSI MI level 2 or 1 in case a single hit is retrieved and no other structures are possible). Furthermore, 1D-<sup>1</sup>H-NMR can

define the type of sugars attached to the flavonoid and mostly also how they are linked.

- v) The complete structural elucidation and confirmation can then be performed with all acquired spectral information in combination with other possible information sources (i.e. MSI MI level 1). Additional 2D-NMR experiments could be needed for full elucidation, especially in case the metabolite to be identified is novel and thus not yet present in any compound library. Moreover, fragment assignments and NMR signal assignments can be checked for annotation purposes.

#### *Quantification of metabolites.*

The quantification of identified metabolites is another important aspect of metabolomics studies. As the ionization efficiency strongly varies between metabolites depending upon their exact chemical structure and their biological matrix, metabolite quantification in LC-MS is normally performed by using a calibration series of authentic standards and by comparing the signal intensities found in real samples with the calibration curve obtained. The lack of easily accessible authentic standards of the enormous amount of primary and especially secondary metabolites and their conjugates present in nature makes this a hard challenge.<sup>36</sup> In some studies the lack of an authentic standard has been circumvented by using structure analogues like deconjugated structures; however it has been shown that the MS response factors can be significantly different and in most cases corrections should be applied.<sup>37</sup> NMR is a quantitative approach, meaning that the integrals of the NMR peaks can be directly related to the amount of compound present, by using any reference compound. However, in complex mixtures it can be difficult to find metabolite-specific signals that are not overlapping with those from other metabolites, which also prevents proper quantification. During this PhD study, the use of NMR in quantification of metabolites isolated from complex biological extracts will be examined.

*Automation of metabolite annotation and identification: computational aid.*

Once metabolites are structurally elucidated, their annotated spectral features are normally stored in compound libraries. The currently available metabolomics workflows generate an increasing amount of large data rich MS and NMR profiles, putting more emphasis on extracting relevant information for metabolite annotation and identification from these complex data sets, as well as candidate selection and rejection of partially characterized and still unknown metabolites. The complex behavior of molecules (and in case of fragmentation in the MS, their fragment ions) in the different mass spectrometers and in the NMR spectrometer still poses challenges in terms of predicting the spectral data from molecular structures, and, even more important, vice versa. Here, computer-assisted annotation and identification (C-AID) of metabolites plays a significant role.

In MS based C-AID, the complex behavior of ionized molecules in the gas phase hinders fast development of ESI based fragmentation rules that can be used for de novo structural elucidation of unknown metabolites. Two approaches can be adopted to overcome this issue: i) in silico fragmentation of metabolites based on generic fragmentation rules, and ii) the generation of a database filled with MS<sup>2</sup> or MS<sup>n</sup> spectra of an ever growing amount of metabolites from different chemical classes (like the NIST library for GC-MS)<sup>14</sup>. The first approach has the advantages that the generated fragments can often immediately be rationalized and that generic fragmentation rules can in principle be applied to any kind of structure; however, complex fragmentation behavior (i.e., rearrangements in case of polyphenols) are difficult to model. The advantages of the second approach are that the generated databases are very useful if searched with similar compounds (i.e. metabolites extracted from similar sources) and that the databases can be queried by similarity of the fragmentation spectra without the need for rationalization of fragments; however, these databases will not generate any useful hit for 'structural outliers'. In view of this, it is expected that partially characterized metabolites can be easier discriminated and annotated in the nearby future using semi-automated MS based approaches; however, the *de novo* identifications remain difficult solely based on MS and MS<sup>n</sup> data, and still

need a combination of analytical data (i.e. LC retention times, NMR shift values) in order to solve the complete chemical structure. Three projects within the MI theme of the NMC are dedicated to develop tools for handling MS<sup>n</sup> and NMR spectral data sets and using them for candidate structure generation. Automatic annotation of all metabolites of interest in complex samples is the final end-goal of the metabolite identification pipeline.

*Aim of this PhD project.*

The overall aim of this thesis is to improve current methods for metabolite identification by developing combined HPLC-based mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) platforms that generate robust spectral data. These data will be used to populate MS<sup>n</sup> and NMR databases and will provide structural information of the detected compounds. Subsequently, it is aimed for that the systematic (and preferably semi- or even fully automated) analysis of these spectral data will lead to the complete structural elucidation of the metabolites involved.

*Overview of my contribution to the thesis chapters.*

Chapter 2 describes how I set up a high-mass resolution tandem mass spectrometry (MS<sup>n</sup>) fragmentation method (the so-called spectral tree approach) on an Ion trap - Orbitrap hybrid mass spectrometer (Thermo Scientific), in collaboration with Piotr Kasper and Miguel Rojas-Cherto from Leiden University and Bert Schipper from Plant Research International, Wageningen. Therefore, I tested different settings of the NanoMate (Advion) injector robot and the hybrid mass spectrometer to come to reproducible data sets. Subsequently, I acquired MS<sup>n</sup> data of a series of 121 polyphenolic molecules to test this approach for differentiation and identification of metabolites based on their unique spectral trees. In chapter 3, I tested the annotation power of this spectral tree approach during LC-MS analysis of extracts from *Lycopersicum esculentum* (tomato) and *Arabidopsis thaliana*. Here, I extensively compared the spectral trees obtained online and offline and in both the Ion trap and Orbitrap mass analyzers and was able to

annotate a large series of metabolites using these MS<sup>n</sup> approaches. Chapter 4 describes how LC-MS (using a Bruker MicroTOF-MS) was connected to a SPE-NMR platform, resulting in an (semi-)automated LC-MS-SPE-NMR system that can trap metabolites from a complex tomato extract for obtaining NMR spectra of metabolites using only minute amounts of sample. In collaboration with Velitchka Mihaleva, I applied NMR predictions using the PERCH NMR software for analyses of the acquired 1D-<sup>1</sup>H-NMR spectra of trapped compounds. Chapter 5 illustrates how I used the newly developed platforms for the selection, annotation, and identification of phenolic compounds present in *Camellia sinensis* (tea plant) extracts. By combining the structural information from MS<sup>n</sup> and 1D-<sup>1</sup>H-NMR, 38 phenolic compounds, including complex acylated conjugates of kaempferol and quercetin were identified based on only microgram amounts of compounds. In chapter 6, I slightly adapted the analytical platforms to annotate and identify phenolic conjugates present in human urine after tea intake, as well as several other urinary metabolites. Here, I extensively studied the resulting MS<sup>n</sup> spectral trees and the NMR spectra to annotate 138 metabolites and completely identify 35 metabolites including conjugated breakdown-products from the tea polyphenols. In collaboration with Velitchka Mihaleva, NMR predictions of several glucuronides and sulphates were performed in order to confirm the correct NMR peak assignments based on 1D-<sup>1</sup>H-NMR data only. In addition, 26 hours quantitative excretion profiles were determined for some valerolactone conjugates. In chapter 7, I discuss the current state of metabolite identification in metabolomics studies and overview the major expected challenges for the nearby future.

## References

1. R. Hall, M. Beale, O. Fiehn, N. Hardy, L. Sumner and R. Bino, *Plant Cell*, 2002, **14**, 1437-1440.
2. H. Schroeter, C. Heiss, J. Balzer, P. Kleinbongard, C. L. Keen, N. K. Hollenberg, H. Sies, C. Kwik-Urbe, H. H. Schmitz and M. Kelm, *P.N.A.S.*, 2006, **103**, 1024-1029.
3. C. Manach and J. L. Donovan, *Free Rad. Res.*, 2004, **38**, 771-785.
4. C. Manach, G. Williamson, C. Morand, A. Scalbert and C. Rémésy, *Am. J. Clinic. Nutr.*, 2005, **81**, 230S-242S.
5. T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, A. Bairoch and J. J. M. Bergeron, *Nat. Meth.*, 2010, **7**, 681-685.
6. K. Saito and F. Matsuda, *Ann. Rev. Plant Biol.*, 2010, **61**, 463-489.
7. T. Kind and O. Fiehn, *Bmc Bioinf.*, 2007, **8**, art. no. 105.
8. M. Arita, <http://www.metabolomics.jp>, 2011.
9. J. B. Harborne and C. A. Williams, *Phytochemistry*, 2000, **55**, 481-504.
10. J. Peterson, J. Dwyer, S. Bhagwat, D. Haytowitz, J. Holden, A. L. Eldridge, G. Beecher and J. Aladesanmi, *J. Food Comp. Anal.*, 2005, **18**, 487-501.
11. B. Álvarez-Sánchez, F. Priego-Capote and M. D. L. d. Castro, *Tr. Anal. Chem.*, 2010, **29**, 120-127.
12. C. H. R. deVos, S. Moco, A. Lommen, J. J. Keurentjes, R. J. Bino and R. D. Hall, *Nat. Prot.*, 2007, **2**, 778-791.
13. G. Theodoridis, H. G. Gika and I. D. Wilson, *Tr. Anal. Chem.*, 2008, **27**, 251-260.
14. NIST, <http://www.nist.gov>, 2012.
15. W. B. Dunn, *Phys. Biol.*, 2008, **5**, art. no. 011001.
16. J. W. Allwood and R. Goodacre, *Phytochem. Anal.*, 2010, **21**, 33-47.
17. K. Dettmer, P. A. Aronov and B. D. Hammock, *Mass Spectrom. Rev.*, 2007, **26**, 51-78.
18. S. Moco, R. J. Bino, R. C. H. De Vos and J. Vervoort, *Tr. Anal. Chem.*, 2007, **26**, 855-866.
19. E. Werner, J. F. Heilier, C. Ducruix, E. Ezan, C. Junot and J. C. Tabet, *J. Chromatogr., B*, 2008, **871**, 143-163.
20. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23-60.
21. T. Kind and O. Fiehn, *Bmc Bioinformatics*, 2006, **7**, -.
22. M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-Mcintyre, P. Begley, K. Carroll, D. Broadhurst, A. Tseng, N. Swainston, I. Spasic, R. Goodacre and D. B. Kell, *Analyst*, 2009, **134**, 1322-1332.
23. A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning, *Anal. Chem.*, 2006, **78**, 2113-2120.
24. A. Makarov and M. Scigelova, *J. Chromatogr., A*, 2010, **1217**, 3938-3945.
25. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and C. H. R. de Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.

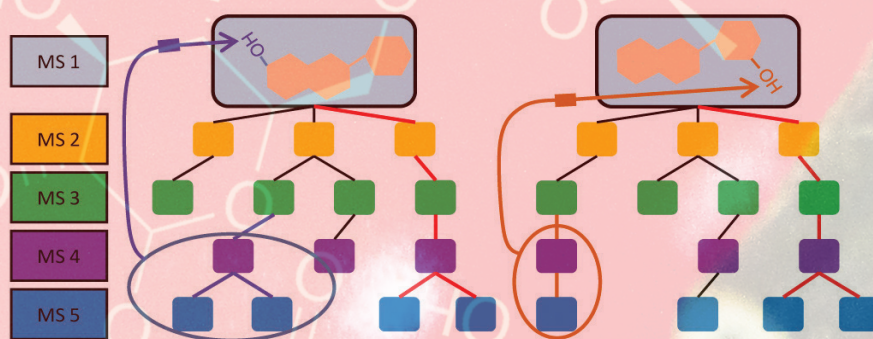
26. M. T. Sheldon, R. Mistrik and T. R. Croleya, *J. Am. Mass Spectr. Soc.*, 2009, **20**, 370-376.
27. M. Spraul, A. S. Freund, R. E. Nast, R. S. Withers, W. E. Maas and O. Corcoran, *Anal. Chem.*, 2003, **75**, 1536-1541.
28. J. W. Jaroszewski, *Planta Med.*, 2005, **71**, 795-802.
29. J. W. Jaroszewski, *Planta Med.*, 2005, **71**, 691-700.
30. C. Clarkson, D. Staerk, S. H. Hansen, P. J. Smith and J. W. Jaroszewski, *J. Nat. Prod.*, 2006, **69**, 1280-1288.
31. C. Clarkson, D. Staerk, S. Honoré Hansen and J. W. Jaroszewski, *Anal. Chem.*, 2005, **77**, 3547-3553.
32. F. Gillotin, P. Chiap, M. Frédérich, J. C. Van Heugen, P. Francotte, P. Lebrun, B. Pirotte and P. De Tullio, *Drug Metab. Dispos.*, 2009, **38**, 232-240.
33. C. Clarkson, M. Sibumb, R. Mensenb and J. W. Jaroszewski, *J. Chromatogr., A*, 2007, **1165**, 1-9.
34. T. Tohge and A. R. Fernie, *Phytochem.*, 2009, **70**, 450-456.
35. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211-221.
36. F. Puiggròs, R. Solà, C. Bladé, M. J. Salvadó and L. Arola, *J. Chromatogr., A*, 2011, **1218**, 7399-7414.
37. T. Farrell, L. Poquet, F. Dionisi, D. Barron and G. Williamson, *J. Pharmaceut. Biomed. Anal.*, 2011, **55**, 1245-1254.







## Chapter 2: Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation



Justin J.J. van der Hooft, Jacques Vervoort, Raoul J. Bino,  
Jules Beekwilder and Ric C.H. de Vos

This chapter was published as research article in *Analytical Chemistry*, 2011, Volume 83 (1), pp 409–416, DOI: 10.1021/ac102546x

### *Abstract*

High mass resolution multistage MS<sup>n</sup> fragmentation was tested for differentiation and identification of metabolites, using a series of 121 polyphenolic molecules. The MS<sup>n</sup> fragmentation approach is based on the systematic breakdown of compounds, forming a so-called spectral tree. A chip-based nano-electrospray ionization source was used combined with an Ion trap, providing reproducible fragmentation, and accurate mass read-out in an Orbitrap Fourier Transform (FT)-MS enabling rapid assignment of elemental formulae to the molecular ions and all fragment ions derived thereof. The used protocol resulted in reproducible MS<sup>n</sup> fragmentation trees up to MS<sup>5</sup>. Obtained results were stable over a 5 month time period, a concentration change of 100-fold, and small changes in normalized collision energy, which is key to metabolite annotation and helpful in structure and substructure elucidation. Differences in the hydroxylation and methoxylation patterns of polyphenolic core structures were found to be reflected by the differential fragmentation of the entire molecule, while variation in a glycosylation site displayed reproducible differences in the relative intensities of fragments originating from the same aglycone fragment ion. Accurate MS<sup>n</sup>-based spectral tree data are therefore a powerful tool to distinguish metabolites with similar elemental formula, thereby assisting compound identification in complex biological samples such as crude plant extracts.

Keywords: Flavonoids, Fragmentation, Isomers, MS<sup>n</sup>, Orbitrap, Spectral Tree

## *Introduction*

One of the bottlenecks in liquid chromatography-mass spectrometry (LC-MS)-based metabolomics is the annotation of the observed molecular ion peaks in an extract. To enable fast elemental formula calculation of detected ions, accurate mass spectrometers, such as time-of-flight (TOF) instruments, Fourier transform ion cyclotron mass spectrometers (FTICR-MS), and Orbitrap Fourier transform (FTMS) machines are used.<sup>1</sup> Nevertheless, the lack of additional structural information frequently hampers translation of the detected accurate mass into its exact chemical structure.<sup>2</sup> For instance, the elemental formula of the flavonoid rutin, C<sub>27</sub>H<sub>30</sub>O<sub>16</sub>, a natural compound commonly present in plants, retrieves 246 different published compound hits in Scifinder (<https://scifinder.cas.org/scifinder>).

MS fragmentation experiments are used to obtain more structural information of detected molecules. Different type of MS instruments, using either nominal or accurate mass detection, are available to fragment ionized molecules and to detect their charged ion fragments.<sup>3</sup> In the tandem-MS technique, MS<sup>2</sup> fragments of the most intense ion or a selected ion are generated. With the use of this approach, several MS/MS databases have been generated in order to facilitate metabolite identification.<sup>4-6</sup> Compared to MS/MS generating MS<sup>2</sup> fragments, so-called MS<sup>n</sup> approaches, in which ions are specifically selected for sequential fragmentation, result in deeper and more detailed fragmentation pathways, thus enabling more structural information. Up to 5 - 8 sequential fragmentation spectra can be obtained, depending upon the concentration and ionization efficiency of the compound. With the use of such a multiple-stage mass spectra approach, we generated so-called spectral trees<sup>7</sup> (Supplemental Figure 1 in the Supporting Information). So far, most MS<sup>n</sup> studies used nominal mass read-out of MS fragments. However, high resolution accurate mass read-out facilitates rapid assignment of an elemental formula to each MS fragment, and thereby to the parent molecule and its fragmentation characteristics.

Mass Spectrometry-based metabolomics approaches are important tools in plant sciences.<sup>8-11</sup> A biologically and economically important class of plant metabolites

is the flavonoids. Flavonoids are polyphenolic compounds ubiquitously present in plant and involved in numerous physiological processes like protection against damage caused by UV light or microbes, and coloring of flowers and fruits.<sup>12</sup> Chemically, the core structure of flavonoids consists of two aromatic and one heteronuclear ring. Flavonoids are divided into several subclasses giving rise to various isomeric aglycones.<sup>13</sup> In plants, flavonoid aglycones normally contain several hydroxyl groups on their aromatic rings. Moreover, one or more sugar moieties can be attached to each of these hydroxyl groups. These and other modifications of the basic flavonoid core have led to a large array of more than 6000 flavonoid species including many different isomers, i.e., structures with similar elemental formulae but different exact chemical configurations. In fact, the large number of different molecules that are possible for a specific elemental formula hampers a fast and unambiguous identification of compounds in biological extracts using MS-based metabolomics<sup>2</sup>. Even in the case of extensive LC separation, in which retention time and absorbance spectra of compounds can be used for identification guidance,<sup>14</sup> the number of isomers having comparable chromatographic behavior and absorbance characteristics prevent an unambiguous annotation based on the elemental formula only.<sup>15</sup> With the use of tandem-MS technologies, MS/MS spectra of a range of flavonoid structures have been investigated and compared.<sup>14, 16</sup> Abad-Garcia et al. recently fragmented a large series of 72 flavonoids with a triple quad MS and derived a number of fragmentation rules based on MS/MS data in positive ionization mode.<sup>14</sup> Likewise, with the use of Ion trap technologies, MS<sup>n</sup> spectra of a number of flavonoids have been generated.<sup>17, 18</sup> Indeed, specific fragments and (sequential) fragment losses related to specific core structures have been reported using MS/MS and MS<sup>n</sup>.<sup>12, 19-22</sup> The fragments from flavonoids originate from two main fragmentation events.<sup>19, 20, 22-30</sup> Ring opening (RO) is characterized by the (subsequent) loss of small neutral molecules like CO, CO<sub>2</sub>, and H<sub>2</sub>O. The other main fragmentation event results in two broken bonds, the so-called cross ring cleavage (CRC) fragmentation, and for polyphenols mostly Retro Dies-Alders (RDA) reactions.

Despite the large number of flavonoid fragmentation studies, data on specific fragmentation patterns that enable discrimination and identification of a range of polyphenolic compounds are scarce.<sup>14, 21, 31</sup> Either in-depth MS<sup>n</sup> fragmentation, with nominal mass read-out, was applied to a small number of flavonoids,<sup>32, 33</sup> or in case of the large study of 72 flavonoids only MS<sup>2</sup> was used.<sup>14</sup> In addition, these studies do not provide data on the reproducibility of the obtained fragmentation spectra and thereby on the robustness of the approach used.

In this study we validated and applied an accurate mass MS<sup>n</sup> spectral tree approach, and we focused on the possibility to discriminate between positional- and stereoisomeric forms. We used a Nanomate injection system coupled to an Ion trap - Orbitrap FTMS system. The NanoMate enabled reproducible and small sample volume infusion, through chip-based nano-electrospray, into the Ion trap, which subsequently fragmented the molecular ion by normalized collision induced dissociation (CID). The most intense ions in the MS<sup>2</sup>, MS<sup>3</sup>, and MS<sup>4</sup> fragmentation spectra were further fragmented while the Orbitrap FTMS provided accurate masses of all ion fragments, enabling fast assignment of their elemental formulas. As an example, a large series of flavonoids with increasing number and size of substitutions was chosen for comparison of their fragmentation spectra in both positive and negative ionization mode (Table S1 in the Supporting Information). The method was validated by acquiring MS<sup>n</sup> data under different experimental conditions and over different time periods. Following this approach, we aimed to improve the discriminative power of MS by generating reproducible fragment ion spectra and MS<sup>n</sup> spectral trees.

### *Materials and Methods*

*Chemicals:* Phenolic standards were obtained from Apin (Oxon, UK), Extrasynthese (Genay, France), Sigma (St. Louis, USA), Fluka (Dorset, UK), and Acros (Geel, Belgium). The purity of all compounds was more than 98%. Quercetin-glucosides were produced in-house through bacterial expression of the tomato UDP-glucose flavonoid glycosyl transferase enzyme. The purified protein was incubated with quercetin aglycone and UDP-glucose as substrates



for 24 hours, allowing the production of a series of mono-, di- and triglucosides. These quercetin-glucosides were separated by preparative LC and their structures determined using Q-TOF accurate mass MS and 2D-NMR. HPLC grade solvents were obtained from Biosolve (Valkenswaard, The Netherlands) and Merck-Schuchardt (Hohenbrunn, Germany). Ultrapure water was made in purification units present in-house.

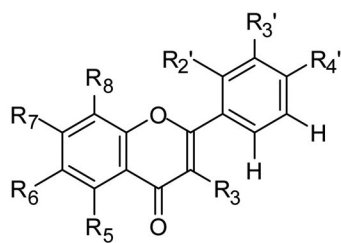
*Preparation of flavonoid standard solutions:* From each polyphenolic molecule, a stock solution of 2.5  $\mu\text{g/mL}$  in 75% MeOH in  $\text{H}_2\text{O}$  acidified with 0.1% formic acid was prepared and stored at  $-20\text{ }^\circ\text{C}$ . Working solutions were prepared by filling 96 wells plates (Abgene) with 40  $\mu\text{L}$  of each stock solution, after which the plates were sealed with thermo foil.

*Mass spectrometry and data handling:* A chip-based nano-electrospray ionization source (Triversa NanoMate, Advion BioSciences) was used for automated direct sample infusion into a LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific) used in negative and positive ionization mode. For each unique  $m/z$  value a separate Xcalibur method was prepared. The LTQ was programmed to use a window of 10 D to isolate the mass of interest in  $\text{MS}^1$ . The data-dependent fragmentation was set as follows:  $\text{MS}^2$  fragmentation of most intense ion in  $\text{MS}^1$ ;  $\text{MS}^3$  fragmentation of the 5 most intense fragment ions in  $\text{MS}^2$ ;  $\text{MS}^4$  fragmentation of the 5 most intense fragment ions in each  $\text{MS}^3$ ;  $\text{MS}^5$  fragmentation of the 3 most intense fragment ions in each  $\text{MS}^4$ . The spectrum list with accurate  $m/z$  values was used to export the  $\text{MS}^n$  spectral tree fragmentation data from Xcalibur into Excel where it was processed. Details on instrument settings and data processing can be found in the Supporting Information.

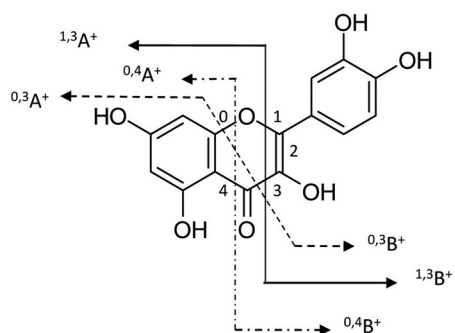
## Results

The chemical structure-spectral tree relationship was studied by individually measuring a series of 121 polyphenolic compounds belonging to different chemical subclasses of flavonoids including various isomeric forms (Table S1 in the Supporting Information), of which a selection is discussed in this article





Flavone basic core



RDA fragmentation through C-ring, shown on quercetin molecule as example

ID	Common name	ID	Common name
1	6,4'-dihydroxyflavone	16	3,7-dimethoxyflavonol
2	6,2'-dihydroxyflavone	17	5,7-dimethoxyflavone
3	6,7-dihydroxyflavone	18	7,8-dimethoxyflavone
4	2'-hydroxyflavonol*	19	3,2'-dimethoxyflavonol
5	7-hydroxyflavonol	20	3,4'-dimethoxyflavonol
6	4'-hydroxyflavonol	21	5,2'-dimethoxyflavone
7	Morin (5,7,2',4'-tetrahydroxyflavonol)	22	quercetin-4'-O-glucoside
8	Quercetin (5,7,3',4'-tetrahydroxyflavonol)	23	quercetin-3-O-glucoside
9	3-methoxyflavonol	24	quercetin-7-O-glucoside
10	5-methoxyflavone	25	quercetin-3,7-O-diglucoside
11	6-methoxyflavone	26	quercetin-3,4'-O-diglucoside
12	7-methoxyflavone	27	quercetin-3,7,3'-O-triglucoside
13	2'-methoxyflavone	28	quercetin-3,7,4'-O-triglucoside
14	4'-methoxyflavone	29	syringetin-3-O-glucoside (5,7,3'-trihydroxy-2',4'-dimethoxyflavonol-3-O-glucoside)
15	3,6-dimethoxyflavonol	30	syringetin-3-O-galactoside (5,7,3'-trihydroxy-2',4'-dimethoxyflavonol-3-O-galactoside)

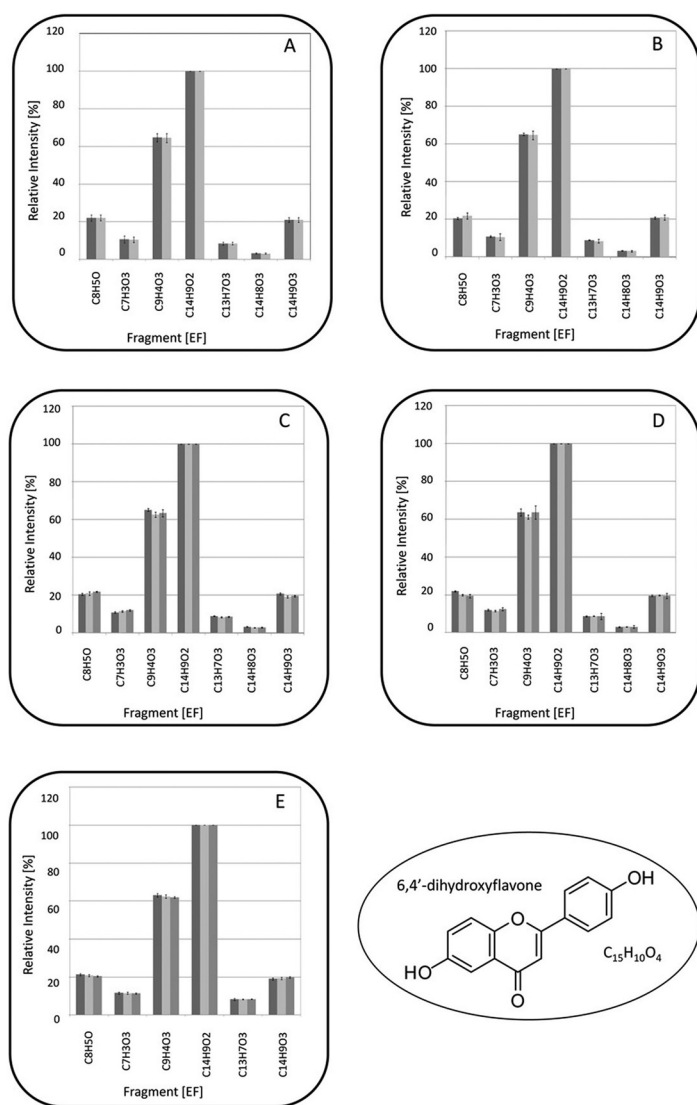
**Figure 1:** The flavone backbone, RDA fragmentation paths, and structures of the discussed polyphenols. The flavone core is displayed at the top left, and Retro Dies-Alders (RDA) fragmentations in the flavone backbone are indicated in the top right with their nomenclature for positive ionization mode. \*A flavonol refers to a 3-hydroxyflavone.

(Figure 1). The resulting spectra were manually checked and processed. Because of the nature of the fragmentation experiments, it was not possible to use a lock mass for automatic recalibration of each scan. Nevertheless, in both ionization modes, the mass deviation was generally less than 1 ppm for MS<sup>1</sup> scans and less than 3 ppm for MS<sup>2-5</sup> scans. In the rare cases that the mass deviation was larger, the mass variation within each scan was always within 2 ppm.

*MS<sup>n</sup> spectral tree method is highly reproducible.*

*Variations over time in fragment intensities:* During the acquisition time, usually multiple (sometimes up to 20) repetitions of the complete fragmentation tree from the same parent ion were acquired. These repetitive MS<sup>n</sup> spectra were averaged and the variation in the intensities of detected fragment ions was calculated, in order to determine the reproducibility of fragmentation patterns. Different fragments for negative and positive ionization mode were obtained, but in both modes these patterns were extremely reproducible. Here, we present validation data from negative mode (Fig. 2), positive mode spectra can be found in the Supporting Information, Supplemental Figure 2. The relative intensities of the fragments in the MS<sup>2</sup> spectra of 6,4'-dihydroxyflavone (**1**) upon combining 6 and 10 spectra are shown in Figure 2A. All 7 selected fragments were detected in all repetitive MS<sup>2</sup> scans, and the overall standard deviation of the relative signal intensities of these fragments was less than 1.2%. For example, the fragment C<sub>9</sub>H<sub>5</sub>O<sub>2</sub> was detected with an average intensity of 22.0% ± 0.83 (mean ± SD) using 6 repetitive spectra, and with an intensity of 22.1% ± 0.79 using 10 repetitive scans. The largest variation observed (< 1.2%) was for fragment C<sub>9</sub>H<sub>4</sub>O<sub>3</sub> for which a relative intensity of 64.7% ± 1.12 over 6 repetitions, and 64.5% ± 1.17 over 10 repetitions was found.

We subsequently compared two independent analyses runs of (**1**), i.e. two infusions and subsequent spectral tree generation within the same day (Figure 2B). Both repetitive infusions yielded highly comparable spectra with the largest relative intensity difference of 1.5% for the C<sub>8</sub>H<sub>5</sub>O fragment, while for the other fragment ions the difference between the analyses was less than 0.6%. Reproducibility



**Figure 2:** Reproducibility of MS<sup>n</sup> patterns in negative ionization mode. MS<sup>2</sup> fragmentation spectra of 6,4'-dihydroxyflavone (**1**) acquired at different conditions are shown: A) averaged 6 (dark grey) or 10 times (light grey); B) two repetitive measurements; C) at 0 months (dark grey), 3 months (light grey), and 5 months (grey); D) at 2.5 µg/ml, 0.5 µg/ml (light grey), and 50 ng/ml (grey); E) at a normalized collision energy of 30% (dark grey), 35% (light grey), and 40% (grey). The average of six spectra is shown and the error bars represent a 95% confidence interval.

of spectral tree generation over time was further determined by analyzing 6,4'-dihydroxyflavone (**1**) and 4'-hydroxyflavonol (**6**), over a 5 month storage (-20 °C) period. The fragmentation spectra of 6,4'-dihydroxyflavone were highly similar within this time period, e.g., the relative intensity of the most varying MS<sup>2</sup> fragment, C<sub>9</sub>H<sub>4</sub>O<sub>3</sub>, differed 2.5% (Figure 2C). For the 4'-hydroxyflavonol, three fragments were detected above 3% of the base peak, of which C<sub>13</sub>H<sub>9</sub>O<sub>2</sub> varied the most (5.2%) (Supplemental Figure 2B in the Supporting Information).

*Variations in fragmentation intensities due to changes in compound concentration and collision energy:* To test whether differences in compound concentrations of the parent ion influence the fragmentation patterns, we generated MS<sup>n</sup> spectra of 6,4'-dihydroxyflavone (**1**) at decreasing initial concentrations of the molecule. The resulting fragmentation spectra (Fig. 2D) indicate that the ion intensity in MS<sup>1</sup>, i.e., the concentration of the molecule to be fragmented, did not significantly influence the relative intensities and thus the relative order of fragment abundance. The measurement variability slightly increased upon lowering the MS<sup>1</sup> ion concentration: from 0.93% at  $2.10 \times 10^7$  ions toward 1.8% at  $8.20 \times 10^7$  ions for the most varying fragment C<sub>9</sub>H<sub>4</sub>O<sub>3</sub>.

The effect of variations in the applied collision energy was tested by generation MS<sup>n</sup> spectra at varying normalized collision energies (Fig. 2E and Supplemental Figure 2D-F). Major changes in neither the relative intensities nor the measurement variation were observed. Only a minor effect was visible in the intensities of all fragments, i.e., on average 1% increase upon increasing the normalized collision energy from 30% to 40%. Thus, small variations in collision energy do not hamper the analysis of the MS<sup>n</sup> spectral trees.

*MS<sup>n</sup> fragmentation generates unique flavonoid spectral trees.*

*Flavonoids with different hydroxylation patterns.* All fragments of the C<sub>15</sub>H<sub>10</sub>O<sub>4</sub> isomers numbered **1-6** (Figure 1) present in the MS<sup>2</sup>-MS<sup>5</sup> spectra in either negative ionization (Table S2 in the Supporting Information) or positive ionization mode (Table S3 in the Supporting Information) were analyzed. Some RO fragments were detected for all isomers, like C<sub>14</sub>H<sub>9</sub>O<sub>2</sub> and C<sub>13</sub>H<sub>9</sub>O<sub>2</sub> in negative ionization mode and C<sub>13</sub>H<sub>9</sub> and C<sub>12</sub>H<sub>9</sub> in positive ionization mode. The MS<sup>2</sup> spectra showed at least one unique CRC fragment for each isomer, in both ionization modes. In addition, the relative intensities of the three most intense RO fragments in MS<sup>2</sup> (i.e. C<sub>13</sub>H<sub>11</sub>O<sub>2</sub>, C<sub>14</sub>H<sub>9</sub>O<sub>2</sub> and C<sub>15</sub>H<sub>9</sub>O<sub>3</sub> in positive ionization mode and C<sub>14</sub>H<sub>8</sub>O<sub>2</sub>, C<sub>14</sub>H<sub>9</sub>O<sub>2</sub> and C<sub>14</sub>H<sub>9</sub>O<sub>3</sub> in negative ionization mode) differed between all six isomers.

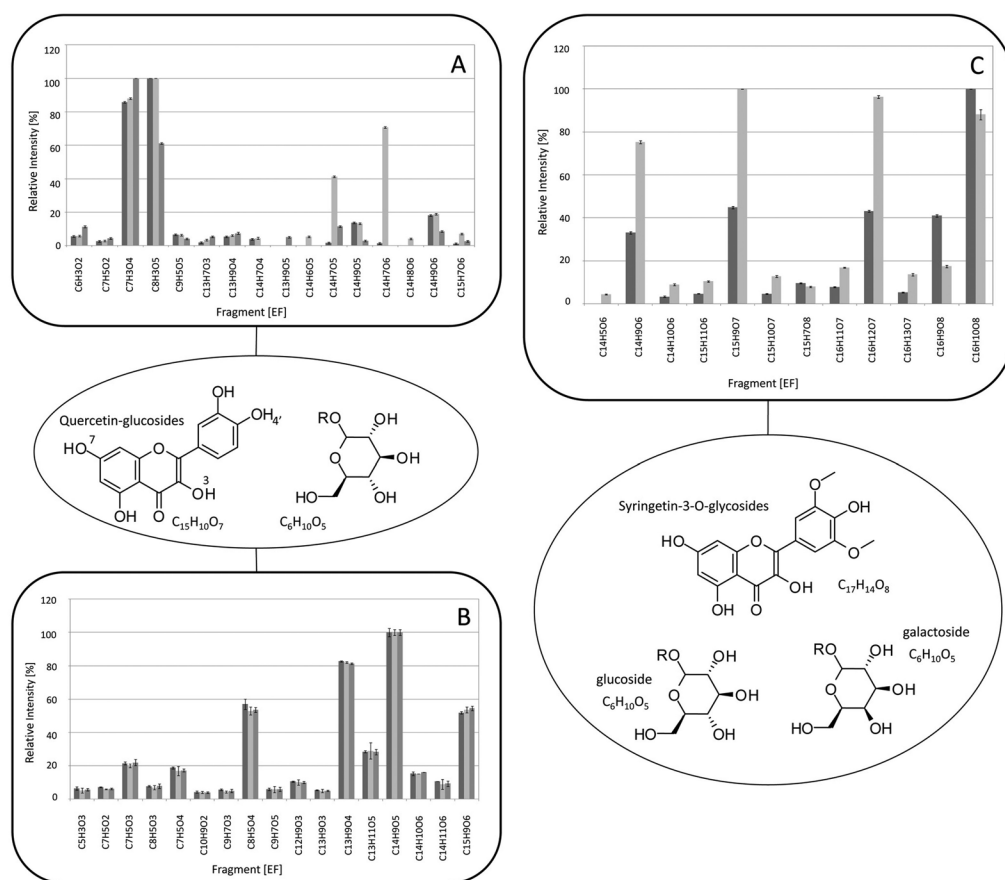
In negative ionization mode both RO and RDA fragmentation yielded radical

fragment ions (Table S2 in the Supporting Information). These radical fragments were sometimes highly abundant in the spectra, for instance the radical C<sub>14</sub>H<sub>8</sub>O<sub>2</sub> was the most intense fragment ion in the MS<sup>2</sup> spectra of (5). Despite being potentially reactive, these radicals showed low measurement variation, e.g., a standard deviation less than 0.59% for C<sub>9</sub>H<sub>4</sub>O<sub>3</sub> of (1).

The two positional isomers of C<sub>15</sub>H<sub>10</sub>O<sub>7</sub>, i.e. morin (7) and quercetin (8), gave differential fragmentation spectra in both ionization modes (Supplemental Figure 3 in the Supporting Information). The different substitution patterns on the B ring resulted in two intense unique CRC fragments in negative ionization mode: C<sub>6</sub>H<sub>5</sub>O<sub>3</sub> for morin and C<sub>8</sub>H<sub>3</sub>O<sub>5</sub> for quercetin. In positive ionization mode, the corresponding morin-specific fragment C<sub>6</sub>H<sub>7</sub>O<sub>3</sub> was detected, though only as a minor fragment, whereas for quercetin the C<sub>8</sub>H<sub>5</sub>O<sub>5</sub> fragment was not present.

*Flavonoids with different methoxylation patterns:* A series of six isomeric monomethoxylated flavonoids (9-14, C<sub>16</sub>H<sub>12</sub>O<sub>3</sub>), and 7 dimethoxylated flavonoids (15-21, C<sub>17</sub>H<sub>15</sub>O<sub>4</sub>) generated fragmentation patterns only in the positive ionization mode, and did not ionize well in negative mode. All fragments collected from the MS<sup>n</sup> spectra of the tested monomethoxylated and dimethoxylated compounds are provided in Tables S4 and S5 in the Supporting Information, respectively.

The monomethoxylated flavonoids produced four fragments that were present in all six isomers, of which the radical fragment C<sub>15</sub>H<sub>10</sub>O<sub>3</sub> was dominant in the MS<sup>2</sup> spectra (Table S4). The RO fragments C<sub>13</sub>H<sub>10</sub>O and C<sub>12</sub>H<sub>9</sub> and the RDA fragment C<sub>6</sub>H<sub>4</sub>O<sub>2</sub> were specific for the A ring mono-substituted methoxyflavones, while C<sub>11</sub>H<sub>10</sub> and C<sub>7</sub>H<sub>5</sub>O<sub>2</sub> were only present in the B ring substituted monomethoxylated flavones. The RO fragment C<sub>14</sub>H<sub>9</sub>O<sub>2</sub> was specific for the 3-methoxyflavonol. All monomethoxylated isomers showed distinct qualitative and/or quantitative differences in their MS<sup>n</sup> fragments. The MS<sup>2</sup> spectra of all dimethoxylated structures were dominated by the radical fragment C<sub>16</sub>H<sub>12</sub>O<sub>4</sub>, due to the loss of one CH<sub>3</sub> radical. Six of the seven isomers produced one or two unique RDA fragments (Table S5 in the Supporting Information).



**Figure 3:** MS<sup>3</sup> spectra of flavonoid glycoside isomers. The MS<sup>3</sup> spectra of the quercetin fragment ions of three quercetin-glycosides (22–24) are shown for A) negative and B) positive ionization mode: 4'-O-glucoside (dark grey), 3-O-glucoside (light grey) and 7-O-glucoside (grey). C): MS<sup>3</sup> fragmentation spectra of the syringetin fragment ions of syringetin-3-O-galactoside (30) (dark grey) and syringetin-3-O-glucoside (29) (light grey) in negative ionization mode. The average of six spectra is shown and the error bars represent a 95% confidence interval.

Flavonoids with different glycosylation patterns: MS<sup>3</sup> spectra in positive mode of all three quercetin-monoglucosides (22–24) tested showed identical fragmentation behavior, resulting from fragmentation of the C<sub>15</sub>H<sub>11</sub>O<sub>7</sub> aglycone fragment (Figure 3B). In contrast, in negative ionization mode the differences between their MS<sup>3</sup> spectra were striking: the ratios of the four most intense fragments were discriminative for all positional isomers (Figure 3A). The RDA fragment C7H4O3 was the base peak for quercetin-7-O-glucoside, while the RO

fragments C<sub>14</sub>H<sub>7</sub>O<sub>5</sub> and C<sub>14</sub>H<sub>7</sub>O<sub>6</sub> were much more intense for quercetin-3-*O*-glucoside (**23**). Some minor fragments were also unique for each glucoside, e.g., C<sub>13</sub>H<sub>9</sub>O<sub>5</sub> for the 7-*O*-glucoside and both C<sub>14</sub>H<sub>6</sub>O<sub>5</sub> and C<sub>14</sub>H<sub>8</sub>O<sub>6</sub> for the 3-*O*-glucoside.

Similarly, the di- and triglucosides of quercetin (**25-28**) did not show differences in their MS<sup>*n*</sup> spectra in positive mode, while in negative mode the intensity patterns of the fragments from the aglycone fragment C<sub>15</sub>H<sub>9</sub>O<sub>7</sub> (in MS<sup>4</sup> and MS<sup>5</sup> for the di- and triglucosides, respectively) differed significantly. Quercetin-3,7,3'-*O*-triglucoside (**27**) and its 3,7,4'-*O*-isomer (**28**) could be discriminated solely on the basis of MS<sup>5</sup> fragmentation of their quercetin aglycone. In the case of quercetin diglucosides, the patterns of the radical fragments in MS<sup>2</sup> and MS<sup>3</sup> were markedly different as well.

To test whether the spectral tree approach would be able to discriminate different hexose sugars attached to the 3-*O* position on the flavonoid aglycone, we compared spectral trees from syringetin-3-*O*-glucose (**29**) and syringetin-3-*O*-galactose (**30**). In line with the quercetin glucosides, positive ionization mode did not yield any differences in the fragmentation. In negative mode, the fragmentation behavior was markedly different between both hexose isomers (Figure 3C).

## *Discussion*

MS/MS and MS<sup>*n*</sup> fragmentation are frequently used as tools in metabolite annotation, based on fragmentation data of reference compounds.<sup>3-7</sup> Here we present an MS<sup>*n*</sup> fragmentation approach using high-resolution Orbitrap FTMS, facilitating elemental formula assignment of all ion fragments and resulting in highly reproducible MS<sup>*n*</sup> fragmentation patterns in both ionization modes. Our method thus enables the detection of even subtle differences in fragmentation patterns. The power of this approach is illustrated by the fact that 119 out of the 121 flavonoids tested, including various isomeric forms, could be discriminated based on the presence of either unique fragments (i.e. higher than 3% of the base peak)

or marked differences (more than about 30%) in the relative intensity of one or more fragments, resulting in unique spectral trees. Only for the isomeric couple catechin and epicatechin the MS<sup>n</sup> fragmentation approach resulted in identical fragment patterns, in both positive and negative ionization mode, which is in line with previous studies.<sup>19, 34, 35</sup> The reproducibility and discriminative power of MS<sup>n</sup> are important factors in the applicability of spectral trees as a tool in the annotation of known metabolites, based on their unique spectral tree, as well as in the identification of unknown metabolites, based on in-depth fragmentation paths providing structural information of the molecule or its sub-structures.

The observed characteristics are essential for the generation of an MS<sup>n</sup> metabolite database. In existing web-based MS/MS databases mostly nominal mass data and MS<sup>2</sup> data are gathered.<sup>36</sup> Lee and co-workers reported the set up of a fragmentation database for flavonoids using Mass Frontier software.<sup>26</sup> However, the Mass Frontier software has some important limitations like inclusion of ghost peaks and noise signals that interfere with the metabolite fragments. Within the Netherlands Metabolomics Centre, new algorithms are under development providing easy processing and analysis of spectra. This will facilitate more global analysis of spectral data, enabling new ways of fragmentation analysis. Therefore, the present study is important in showing that the robust MS<sup>n</sup> spectral tree approach can provide the input for a web-based MS<sup>n</sup> fragmentation database, thereby expanding the possibilities of MS in the metabolite identification process. While the described results are based on pure compounds infused by the Nanomate one by one, the practical use and robustness of MS<sup>n</sup> fragmentation during LC-MS-based metabolome profiling still need to be shown. This application of on-line LC-MS<sup>n</sup> in metabolite annotation of crude extracts is currently under investigation. With identical protocols, the comparison of fragmentation data of LTQ-Orbitrap systems at different labs, as well as fragmentation data from different Ion trap platforms, is facilitated. Currently, a round robin experiment is being conducted in order to examine the intra and inter-laboratory variations of the LTQ-Orbitrap platforms.



*Reproducibility and uniqueness of MS<sup>n</sup> spectral trees:* Parameters that can potentially influence the MS<sup>n</sup> fragmentation patterns, such as normalized collision energy, time interval and compound concentration, only slightly influenced the spectral tree characteristics, from MS<sup>2</sup> up to MS<sup>5</sup>. For example, the relative intensities of the fragments did not change from an ion concentration range of \*10<sup>5</sup> up to \*10<sup>7</sup> counts, i.e., a 100 fold difference. However, in case the discrimination is based on specific low abundant fragments present in lower stage MS<sup>n</sup> spectra, e.g. MS<sup>2</sup> or MS<sup>3</sup>, or on differences in only higher stage MS<sup>n</sup> spectra, e.g. MS<sup>4</sup> or MS<sup>5</sup>, the spectral tree may lose its uniqueness at too low metabolite concentrations. This threshold concentration depends upon both the ionization and fragmentation efficiency of the selected metabolite. In line with Kite and Veitch,<sup>37</sup> we observed that within one measurement series it is very well possible to discriminate between different flavonoid isomers, including the differences in relative intensities of generic fragments. On top of that, we established that the MS<sup>n</sup> spectral trees are also highly comparable between different measurement series even over a 5 month period, despite slight variations.

*Accurate mass MS<sup>n</sup>:* The studies on flavonoid fragmentation so far were hampered by either limited fragmentation depth in MS/MS<sup>14, 26, 28, 38-42</sup> or limited mass accuracy.<sup>17, 18, 32, 43-45</sup> Here we combined accurate mass with in-depth MS<sup>n</sup> fragmentation, using an Ion trap MS coupled to an Orbitrap FTMS. The accurate mass MS<sup>n</sup> approach enabled elemental composition calculation of all detected fragments. The assignment of the MS<sup>2</sup> fragments from 6,7-dihydroxyflavone (**3**) in positive ionization mode underlined the benefit of high mass resolution and accuracy: the fragments C<sub>7</sub>H<sub>5</sub>O<sub>4</sub> and C<sub>12</sub>H<sub>9</sub> are close in exact mass, i.e., 51.6 milliDalton (mDa) difference. In nominal mass read-out, which has so far been used in all Ion trap fragmentation studies on flavonoids<sup>17, 18, 32, 43-45</sup>, the assignment of these two elemental formulas is hampered, thereby preventing their structural elucidation. Compared to accurate mass MS/MS, the present accurate MS<sup>n</sup> approach yielded more informative fragmentation spectra, which in some cases was essential to discriminate flavonoid isomers. For instance, differences between flavonoid glycosides were more prominent in MS<sup>3</sup> than in MS<sup>2</sup>, while

quercetin-triglucosides showed differential spectra only in MS<sup>5</sup>.

So far, most studies have used positive ionization mode for polyphenol identification, because they conclude that negative mode fragmentation often results in more complex and therefore less informative spectra. Indeed, we observed more complex MS<sup>n</sup> fragmentation patterns in negative ionization mode as compared to positive ionization, but nevertheless the fragmentation spectra were stable and reproducible, and therefore representing a fingerprint of the fragmented molecule. Moreover, it was only in negative mode that the various flavonoid glycosides showed differential and specific fragmentations.

The accurate mass MS<sup>n</sup> approach led to a collection of fragments occurring in series of hydroxylated and methoxylated flavones (Suppl. Tables 2-5). Most MS<sup>n</sup> fragments we detected from hydroxylated flavonoids have been described in literature before;<sup>46</sup> however, not all have been described with proposed fragmentation paths. For example, the fragment C<sub>12</sub>H<sub>9</sub> (detected within 2 ppm of its expected mass) was present in the spectra of all dihydroxyflavones in both positive and negative ionization mode. The generation of this fragment ion is remarkable, as apparently all (four) oxygen atoms have been lost from the flavonoid core before generation of the C<sub>12</sub>H<sub>9</sub> ion. We propose a fragmentation pathway from, for instance, the protonated 4'-hydroxyflavonol towards C<sub>12</sub>H<sub>9</sub>, involving sequential losses of three CO and one H<sub>2</sub>O molecules (Supplemental Figure 4 in the Supporting Information). Thus, the accurate mass MS<sup>n</sup> spectral tree method can corroborate proposed pathways and develop new ideas about the fragmentation paths present in polyphenolic compounds, thereby facilitating structural elucidation of molecular ions.

*Generation of radical fragment ions:* The production of radicals in the fragmentation of polyphenolic compounds has been observed and described before both in positive<sup>22, 24, 46</sup> and in negative ionization mode.<sup>22, 32, 37, 46-48</sup> The chemistry behind the origin of these radical ions is poorly understood.<sup>47, 49, 50</sup> We observed similar fragmentation patterns, including patterns upon fragmentation of the radical ions, for compounds studied earlier,<sup>21,32</sup> indicating that these radicals are reproducibly generated and are not artifacts of specific MS platforms.

Especially the methoxylated polyphenols give rise to many radical fragments. The 3 Da selection window used in the MS<sup>n</sup> spectral tree method caused non-radical and radical neighbor ions to be fragmented simultaneously. Nevertheless, these complex and mixed radical and non-radical fragmentation patterns were reproducible and often different for related isomeric compounds.

*MS<sup>n</sup> spectra of flavonoid aglycones:* MS<sup>n</sup> fragmentation of hydroxylated and methoxylated flavones resulted in specific RDA fragments that were discriminative between the various isomeric forms tested, either in one or both ionization modes, which can be rationalized because the number of hydroxyl groups in each ring influences the mass of the fragments resulting from the RDA reaction. In addition, the relative intensities of a number of RO fragments were influenced by the different substitution patterns. Some of these observations can be explained by fragmentation rules earlier defined for Electron Impact (EI) fragmentation, e.g. the facilitated loss of water in the case of ortho-positioned hydroxyl groups like in 6,7-dihydroxyflavone (**3**).<sup>29</sup>

By comparison of the fragmentation of dihydroxylated flavonoids with increasingly hydroxylated flavonoids, it appeared that a higher amount of hydroxyl substitutions resulted in a more branched fragmentation tree, i.e., more intense base peak fragments in MS<sup>2</sup> and MS<sup>3</sup> and a higher number of fragments generating MS<sup>4</sup> and MS<sup>5</sup> spectra. As a result, more time was needed to generate a single tree. The same fragmentation behavior was found for methoxylated flavonoids, indicating that substitutions on the aromatic rings weaken the bond strengths within the aromatic ring. Likely, the exact positions of the hydroxyl and methoxyl groups in the molecular ions influence the bonding energies and charge (de)locations over the ions, which determine their subsequent fragmentation. The flavanols catechin and epicatechin were taken as an example in the comparison of the fragmentation patterns of stereoisomeric flavonoid cores, differing only in the stereochemistry of their C3-hydroxylation. Apparently the axial or equatorial C3-hydroxylation does not result in differential charge density of the flavonoid core. Thus, MS<sup>2</sup> and sometimes MS<sup>3</sup> fragmentation will discriminate most flavonoid aglycones; however, following our approach up to MS<sup>5</sup>, more in-depth

fragmentation paths will be revealed, facilitating assignment of structures to fragments.

*MS<sup>n</sup> spectra of flavonoid glycosides:* For the flavonoid glycosides studied, the MS<sup>n</sup> spectra were generally more discriminative in negative ionization mode than in positive ionization mode. In addition, their ionization is generally better in negative mode. This is in contrast to the flavonoid aglycones, for which positive ionization was the favorable mode, especially for those aglycones that did not ionize at all in negative ionization mode such as (poly)methoxylated flavonoids. MS<sup>n</sup> fragmentation was able to differentiate between the three positional monoglucoside isomers of quercetin (**22-24**) in negative ionization mode (Figure 3), in line with recent MS/MS observations by Geng et al.<sup>51</sup> In addition to the monoglucosides, using MS<sup>n</sup> up to MS<sup>5</sup>, we were also able to differentiate between isomers of di- and three-glucosides of quercetin, thus showing the power of MS<sup>n</sup>. The most likely explanation for the observed differences between fragmentation patterns obtained from the same quercetin fragment ion derived from the various quercetin-glucoside structures is the occurrence of differential charge distributions over the quercetin aglycone fragments. Using the spectral tree approach we were able to distinguish glucose from a galactose moiety (C4 epimers) located at the 3-*O* position on the rather complex flavonol cores syringetin (Figure 3) as well as quercetin. The difference in fragmentation behavior between, on the one hand, the glycosidic stereoisomers and, on the other hand, the isomeric flavonoid couple of catechin and epicatechin is remarkable, because both couples differ in the orientation of only one hydroxyl group. A possible explanation for the difference in fragmentation behavior may be a lower flexibility of the hydroxyl group attached to the flavonoid core compared to a more flexible sugar hydroxyl groups, as well as the larger internal flexibility of the sugar moiety.

## *Conclusions*

Accurate mass MS<sup>n</sup> fragmentation spectra were reproducibly generated enabling differentiation of isomeric compounds such as polyphenols. For all 121 but two compounds tested, the spectral trees showed unique fragments and/or differences in relative intensities of fragment ions. Subtle differences between glucose and galactose conjugations and positions of sugar substitution were observed upon fragmentation of the aglycone fragment, in some cases only apparent in MS<sup>5</sup> spectra. The method developed was robust and highly reproducible within and between analyses series. Thus, we conclude that MS<sup>n</sup> fragmentation patterns can be used for structural elucidation of polyphenol structures and substructures. Although we cannot yet rule out the influence of different LTQ-Orbitrap instruments, both the reproducibility and robustness of our data and the correspondence with flavonoid MS<sup>n</sup> data spectra available in literature suggest that MS<sup>n</sup> databases can be developed to compare and match results from different laboratories.

## *Acknowledgements*

This research was granted by the Netherlands Metabolomics Centre (NMC) and the Centre for BioSystems Genomics (CBSG), which are both part of the Netherlands Genomics Initiative (NGI). Piotr Kasper, Rob Vreeken, Nico Nibbering, Velitchka Mihaleva, and Bert Schipper are thanked for advice and/or technical assistance.

## *Supporting Information Available*

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>. The supplemental figures and tables are also part of this thesis, pages 63-78, and the supplemental Excel file is provided on <http://library.wur.nl/WebQuery/edepot/216854>

### List of supplemental data

Supplemental Figure 1 The concept of the MS<sup>n</sup> spectral tree

Supplemental Figure 2 MS<sup>n</sup> patterns in time and at different concentrations and collision energies

Supplemental Figure 3 Fragmentation spectra of the two positional flavonoid aglycones

Supplemental Figure 4 Proposed fragmentation pathway leading to the C<sub>12</sub>H<sub>9</sub> fragment ion

Supplemental Text 1 Experimental details

Table S-1 The complete list of 121 polyphenolic compounds that were fragmented in positive and negative ionization mode was collected in this Excel file.

Table S-2 The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 dihydroxylated flavones (1-6 in Fig. 1 of article) collected in negative ionization mode.

Table S-3 The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 dihydroxylated flavones (1-6 in Fig. 1 of article) collected in positive ionization mode.

Table S-4 The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 monomethoxylated flavones (9-14 in Fig. 1 of article) collected in positive ionization mode.

Table S-5 The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 7 dimethoxylated flavones (15-21 in Fig. 1 of article) collected in positive ionization mode.

### References

1. P. Kiefer, J.-C. Portais and J. A. Vorholt, *Anal. Biochem.*, 2008, **382**, 94-100.
2. T. Kind and O. Fiehn, *Bmc Bioinf.*, 2007, **8**, art. no. 105.
3. E. Werner, J. F. Heilier, C. Ducruix, E. Ezan, C. Junot and J. C. Tabet, *J. Chromatogr. B*, 2008, **871**, 143-163.
4. F. Matsuda, K. Yonekura-Sakakibara, R. Niida, T. Kuromori, K. Shinozaki and K. Saito, *Plant J.*, 2009, **57**, 555-577.
5. D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. de souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhtudinov, L. Li, H. J. Vogel and I. Forsythei, *Nucleic Acids Res.*, 2009, **37**.
6. H. Horai, M. Arita and T. Nishioka, in *BioMedical Engineering and Informatics: New Development and the Future - Proceedings of the 1st International Conference on BioMedical Engineering and Informatics, BMEI 2008*, Sanya, Hainan, 2008, pp. 853-857.
7. M. T. Sheldon, R. Mistrik and T. R. Croleya, *J. Am. Mass Spectr. Soc.*, 2009, **20**, 370-376.
8. W. B. Dunn, *Phys. Biol.*, 2008, **5**, art. no. 011001.
9. R. Hall, M. Beale, O. Fiehn, N. Hardy, L. Sumner and R. Bino, *Plant Cell*, 2002, **14**, 1437-1440.

10. R. D. Hall, *New Phytol.*, 2006, **169**, 453–468.
11. J. M. Hagel and P. J. Facchini, *Phytochem. Rev.*, 2008, **7**, 479-497.
12. J. B. Harborne, *Comparative Biochemistry of the flavonoids*, Academic Press London and New York, 1967.
13. E. de Rijke, P. Out, W. M. A. Niessen, F. Ariese, C. Gooijer and U. A. T. Brinkman, *J. Chromatogr., A*, 2006, **1112**, 31-63.
14. B. Abad-García, L. A. Berrueta, S. Garmón-Lobato, B. Gallo and F. Vicente, *J. Chromatogr., A*, 2009, **1216**, 5398-5415.
15. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and C. H. R. de Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.
16. R. J. Hughes, T. R. Croley, C. D. Metcalfe and R. E. March, *Int. J. Mass Spectrom.*, 2001, **210-211**, 371-385.
17. M. Yang, W. Wang, J. Sun, Y. Zhao, Y. Liu, H. Liang and D. A. Guo, *Rap. Commun. Mass Spectrom.*, 2007, **21**, 3833-3841.
18. H. Y. Zhao, J. H. Sun, M. X. Fan, L. Fan, L. Zhou, Z. Li, J. Han, B. R. Wang and D. A. Guo, *J. Chrom., A*, 2008, **1190**, 157-181.
19. D. Tsimogiannis, M. Samiotaki, G. Panayotou and V. Oreopoulou, *Molecules*, 2007, **12**, 593-606.
20. N. Fabre, I. Rustan, E. de Hoffmann and J. Quetin-Leclercq, *J. Am. Soc. Mass Spectrom.*, 2001, **12**, 707-715.
21. F. Kuhn, M. Oehme, F. Romero, E. Abou-Mansour and R. Tabacchi, *Rap. Commun. Mass Spectrom.*, 2003, **17**, 1941-1949.
22. R. Liu, M. Ye, H. Guo, K. Bi and D.-a. Guo, *Rap. Commun. Mass Spectrom.*, 2005, **19**, 1557-1565.
23. G. Liu, Z. Xu, J. Chen, G. Lang, Q. Tian, Y. Shen, B. Chen and S. Yao, *J. Chrom., B*, 2009, **877**, 2545-2550.
24. U. Justesen, *J. Mass Spectrom.*, 2001, **36**, 169-178.
25. F. Cuyckens, R. Rozenberg, E. De Hoffmann and M. Claeys, *J. Mass Spectrom.*, 2001, **36**, 1203-1210.
26. J. S. Lee, D. H. Kim, K.-H. Liu, T. K. Oh and C. H. Lee, *Rap. Commun. Mass Spectrom.*, 2005, **19**, 3539-3548.
27. R. E. March, E. G. Lewars, C. J. Stadey, X. S. Miao, X. Zhao and C. D. Metcalfe, *Int. J. Mass Spectrom.*, 2006, **248**, 61-85.
28. F. Cuyckens and M. Claeys, *J. Mass Spectrom.*, 2004, **39**, 1-15.
29. J. T. Watson and O. D. Sparkman, *Introduction to Mass Spectrometry*, Wiley, West Sussex, England, 2007.
30. F. Cuyckens and M. Claeys, *J. Mass Spectrom.*, 2005, **40**, 364-372.
31. M. N. Clifford, S. Knight and N. Kuhnert, *J. Agr. Food Chem.*, 2005, **53**, 3821-3832.
32. J. Kang, L. A. Hick and W. E. Price, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 857-868.
33. J. L. Wolfender, P. Waridel, K. Ndjoko, K. R. Hobby, H. J. Major and K. Hostettmann, *Analisis*, 2000, **28**, 895-906A.
34. G. Shui and L. P. Leong, *J. Chrom., A*, 2004, **1022**, 67-75.
35. J. Wollgast, L. Pallaroni, M. E. Agazzi and E. Anklam, *J. Chrom., A*, 2001, **926**, 211-220.
36. T. Tohge and A. R. Fernie, *Phytochemistry*, 2009, **70**, 450-456.

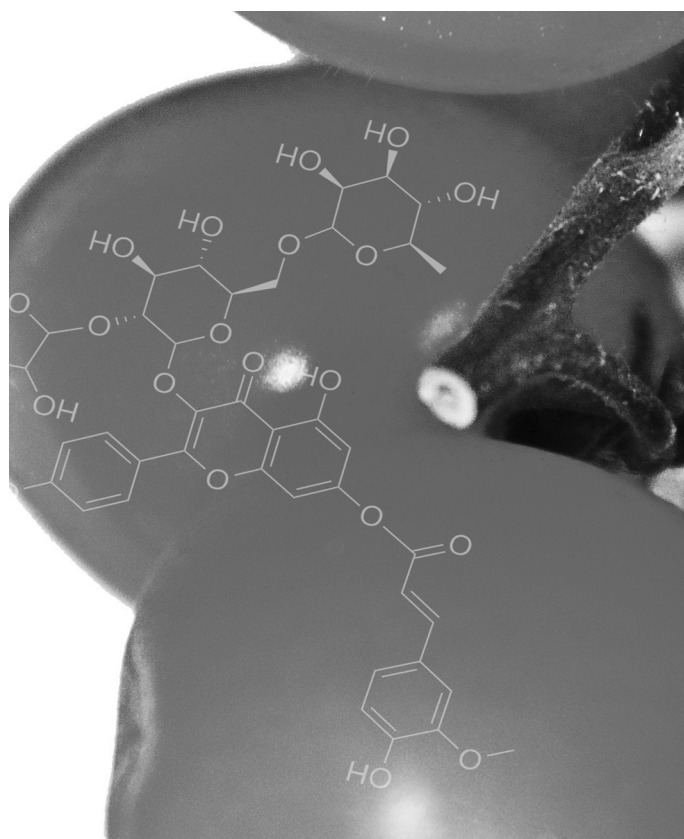
37. E. Hvattum and E. Dag, *J. Mass Spectrom.*, 2003, **38**, 43-49.
38. G. C. Justino, C. M. Borges and H. M. Florêncio, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 237-248.
39. J. L. Wolfender, S. Rodriguez and K. Hostettmann, *J. Chrom., A*, 1998, **794**, 299-316.
40. F. Cuyckens, Y. L. Ma, G. Pocsfalvi and M. Claeys, *Analisis*, 2000, **28**, 888-895.
41. P. Miketova, K. H. Schram, J. Whitney, M. Li, R. Huang, E. Kerns, S. Valcic, B. N. Timmermann, R. Rourick and S. Klohr, *J. Mass Spectrom.*, 2000, **35**, 860-869.
42. M. A. Farag, D. V. Huhman, Z. Lei and L. W. Sumner, *Phytochemistry*, 2007, **68**, 342-354.
43. P. Kachlicki, Å. Marczak, L. Kerhoas, J. Einhorn and M. Stobiecki, *J. Mass Spectrom.*, 2005, **40**, 1088-1103.
44. X. Meng, H. Li, F. Song, C. Liu, Z. Liu and S. Liu, *Chin. J. Chem.*, 2009, **27**, 299-305.
45. G. M. Weisz, D. R. Kammerer and R. Carle, *Food Chem.*, 2009, **115**, 758-765.
46. D. C. Burns, D. A. Ellis, H. Li, E. G. Lewars and R. E. March, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 437-454.
47. E. G. Lewars and R. E. March, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 1669-1679.
48. K. Ablajan, Z. Abliz, X. Shang, J. He, R. Zhang and J. Shi, *J. Mass Spectrom.*, 2006, **41**, 352-360.
49. B. D. Davis and J. S. Brodbelt, *J. Mass Spectrom.*, 2008, **43**, 1045-1052.
50. M. Born, S. Ingemann and N. M. M. Nibbering, *Mass Spectrom. Rev.*, 1997, **16**, 181-200.
51. P. Geng, J. Sun, R. Zhang, J. He and Z. Abliz, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 1519-1524.







Supporting information to *Chapter 2*:  
'Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation'



Justin J.J. van der Hooft, Jacques Vervoort, Raoul J. Bino,  
Jules Beekwilder and Ric C.H. de Vos

This Supporting Information was published online with the research article in *Analytical Chemistry*, 2011, Volume 83 (1), pp 409–416, DOI: 10.1021/ac102546x

*Supplemental text 1: Experimental details*

*NanoMate sample injection procedure:* A chip-based nano-electrospray ionization source (Triversa NanoMate, Advion BioSciences) was used for automated direct sample infusion. Delivery of the solutions to the MS was obtained by programming the NanoMate to firstly pinch a hole into the sealing foil that covered the 96-wells plate with a pipette tip, and then to take 15  $\mu\text{l}$  of solution from the well and to infuse this sample into the MS via a clean spray nozzle of the chip. During analyses, the sample plate was cooled to 10°C. All compounds were subjected to MS<sup>n</sup> fragmentation in both negative and positive ionization modes. The NanoMate parameter settings were as follows: spray voltage 1.6 kV and 0.45 psi N<sub>2</sub> gas for negative ionization mode, and 1.4 kV and 0.40 psi N<sub>2</sub> gas for positive ionization mode. After 30 seconds of infusion, electrospray sensing was switched on. The NanoMate was programmed to take a new nozzle and continue infusion when the spray current dropped below 4 nanoAmpère (nA) or was above 400 nA for more than 90 seconds. Following sample infusion and MS analysis, the pipette tip was ejected and a new tip and nozzle were used for the next sample, thereby preventing any cross-contamination and carryover. One MS<sup>n</sup> experiment consisted of 30 s delay and 15-20 min of recording spectra, depending upon the time of the MS<sup>n</sup> cycle length.

*Mass spectrometry details:* The MS system consisted of a LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific), with Xcalibur software to control instruments and for data acquisition and analysis. Before a measurement series, the Orbitrap was externally calibrated using NaFormate clusters (100-1000 range) and tuned for  $m/z$  384.9353 (negative ionization mode) or  $m/z$  362.9263 (positive ionization mode). The capillary temperature was set to 200 °C. The Ion trap settings used were: AGC target value of 30,000 and 10,000 in full scan and MS<sup>n</sup> mode respectively. The FTMS settings used in the MS1 scan event were: mass resolution 60,000, 2 microscans, and an AGC target value of 200,000 charges. For all dependent scan events, the following settings were used: mass

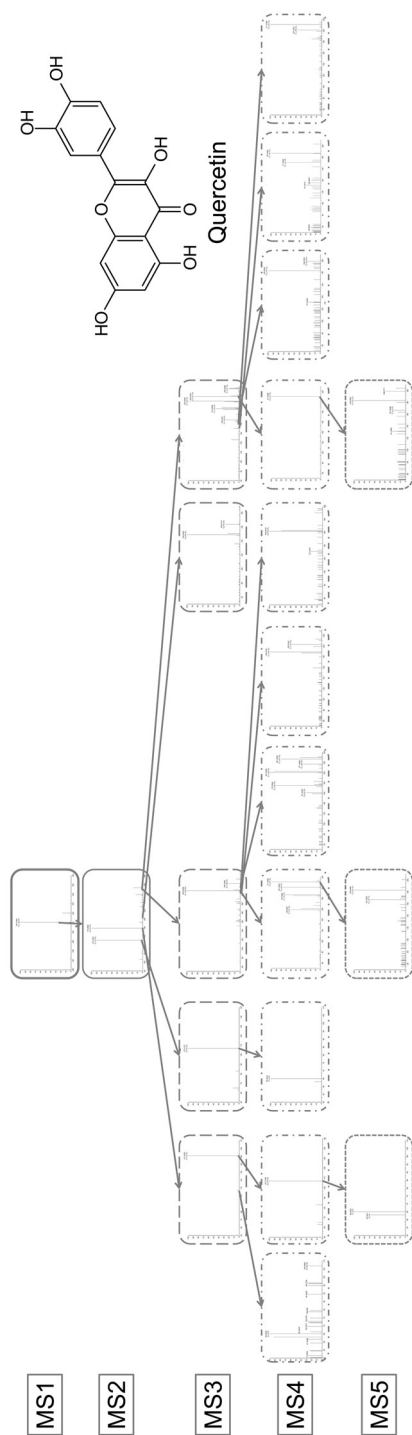
resolution 15,000, 2 microscans, AGC target of 100,000, isolation width 3 Dalton, threshold 50,000 counts/s, normalized collision energy of 35%. The following additional microscans, on top of the 2 microscans, were used: 11 in MS<sup>2</sup>, 21 in MS<sup>3</sup>, 31 in MS<sup>4</sup> and 41 in MS<sup>5</sup>. Additional ion filling of the Ion Trap on top of the 30,000 in the MS1 setting were 100,000 charges for MS<sup>2</sup>, 200,000 charges for MS<sup>3</sup>, 300,000 charges for MS<sup>4</sup>, and 400,000 charges for MS<sup>5</sup>.

For each unique m/z value a separate Xcalibur method was prepared. The LTQ was programmed to use a window of 10 D to isolate the mass of interest in MS1. The data-dependent fragmentation was set as follows: MS<sup>2</sup> fragmentation of most intense ion in MS1; MS<sup>3</sup> fragmentation of the 5 most intense fragment ions in MS<sup>2</sup>; MS<sup>4</sup> fragmentation of the 5 most intense fragment ions in each MS<sup>3</sup>; MS<sup>5</sup> fragmentation of the 3 most intense fragment ions in each MS<sup>4</sup>.

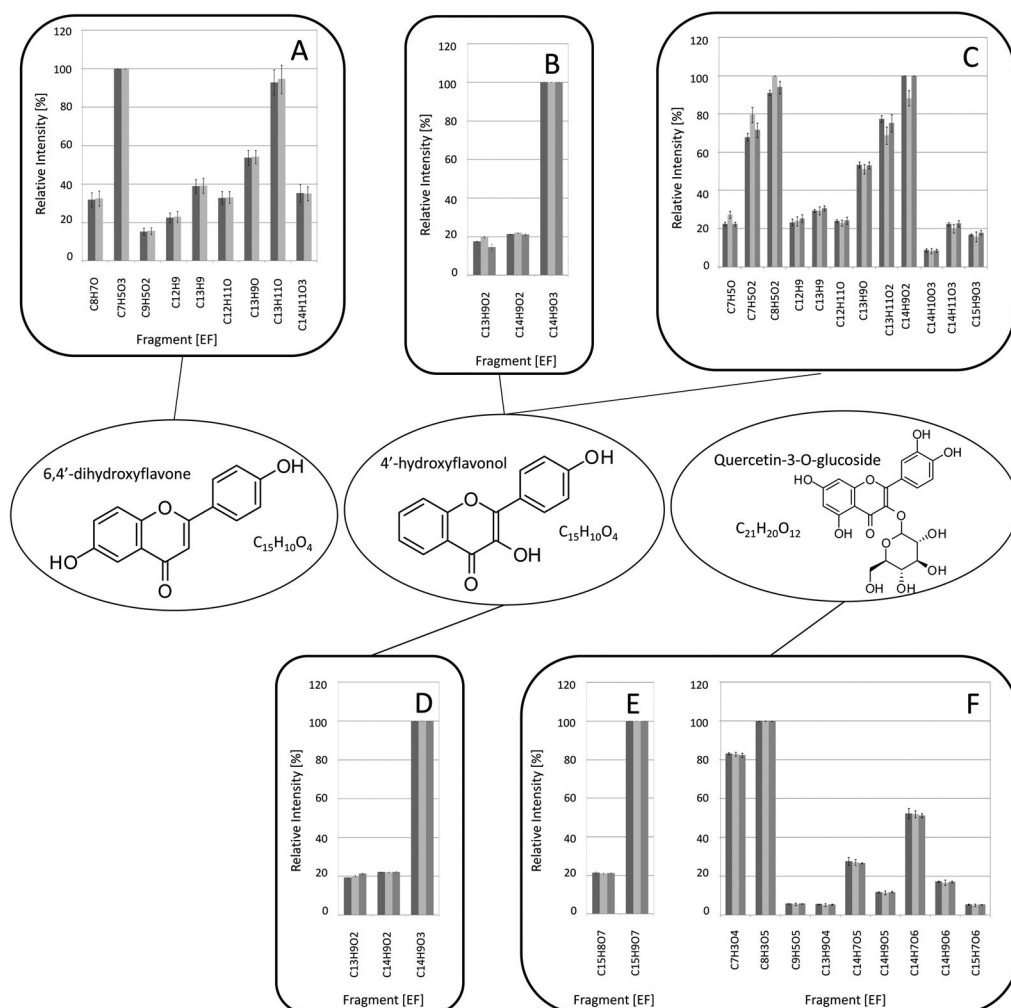
*Data processing:* The settings for automatic elemental formula calculation were as follows: maximum mass deviation: 15 ppm, charge state 1, maximum numbers of possible C, H, and O, were, respectively, 60, 100 and 50. The nitrogen rule was not applied in order to include possible radical ions.

The spectrum list with accurate m/z values was used to export the MS<sup>n</sup> spectral tree fragmentation data from Xcalibur into Excel. The relative intensities of consecutive scans per specific MS<sup>n</sup> event were averaged. If the fragmented ion was the base peak, i.e., the most intense peak in the spectrum, this peak was excluded. A threshold of 3% relative intensity, as compared to the base peak, was used for selecting masses in each fragmentation spectrum. Elemental formulae of fragments were checked on the ring and double bond (RDB) factor for correctness, as well as for any violations with the parental formula.

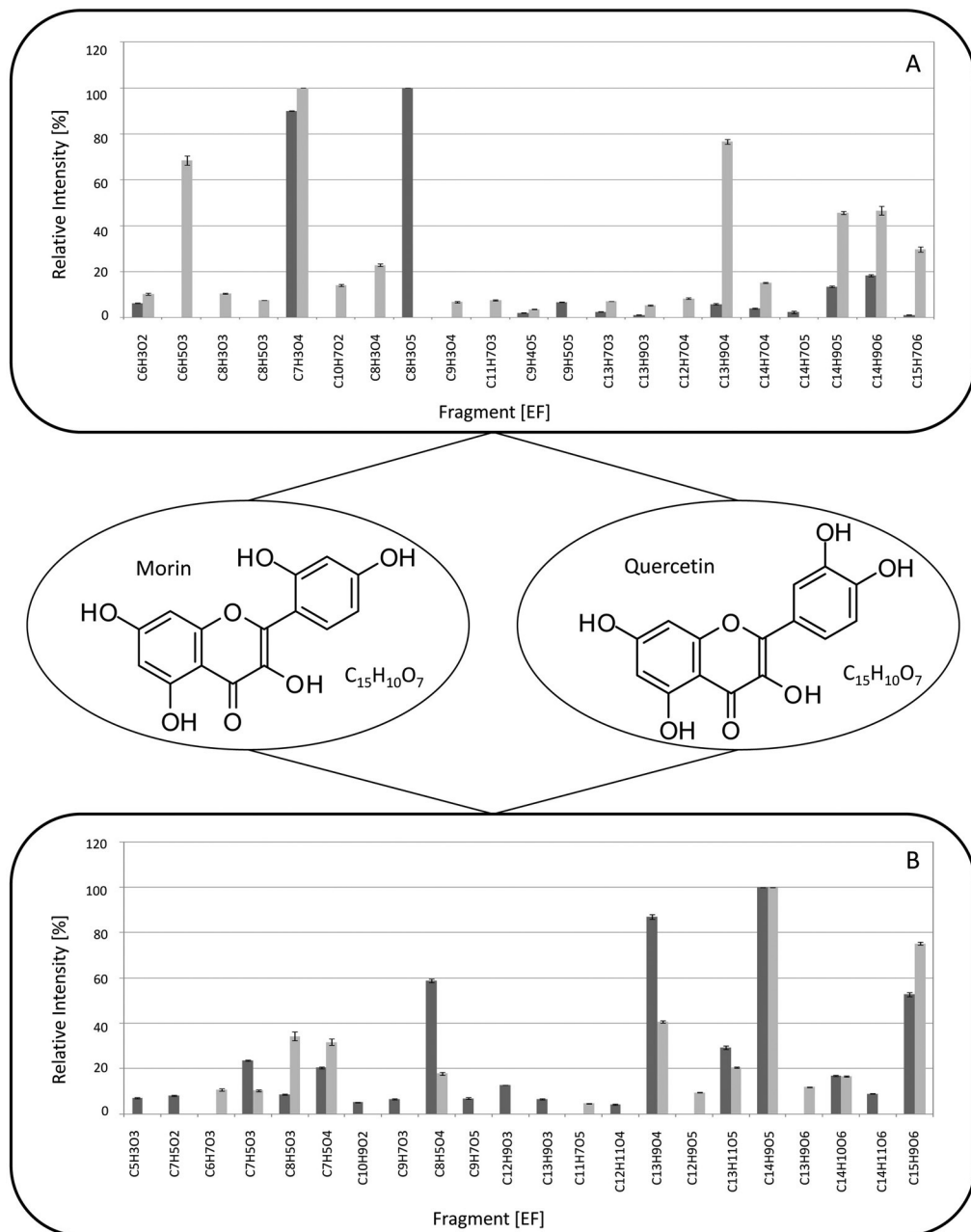
The supplementary tables (S2-S5) containing the fragments present in all MS<sup>n</sup> scans of different sets of isomeric flavonoids were created by manually checking the MS<sup>n</sup> data and confirming the presence of the fragments in 3 spectral tree repetitions. The nomenclature adopted from Ma et al.<sup>34</sup> was used to annotate the RDA derived fragments (Fig. 1).



**Supplemental Figure 1:** The concept of the  $MS^n$  spectral tree is visualized for the fragmentation of quercetin in negative ionization mode. The  $MS^1$  is followed by one  $MS^2$  scan, after which five  $MS^3$  spectra are displayed, generated of the 5 most intense fragments in  $MS^2$ . Several  $MS^4$  spectra and even some  $MS^5$  spectra were obtained for this relatively small molecule.

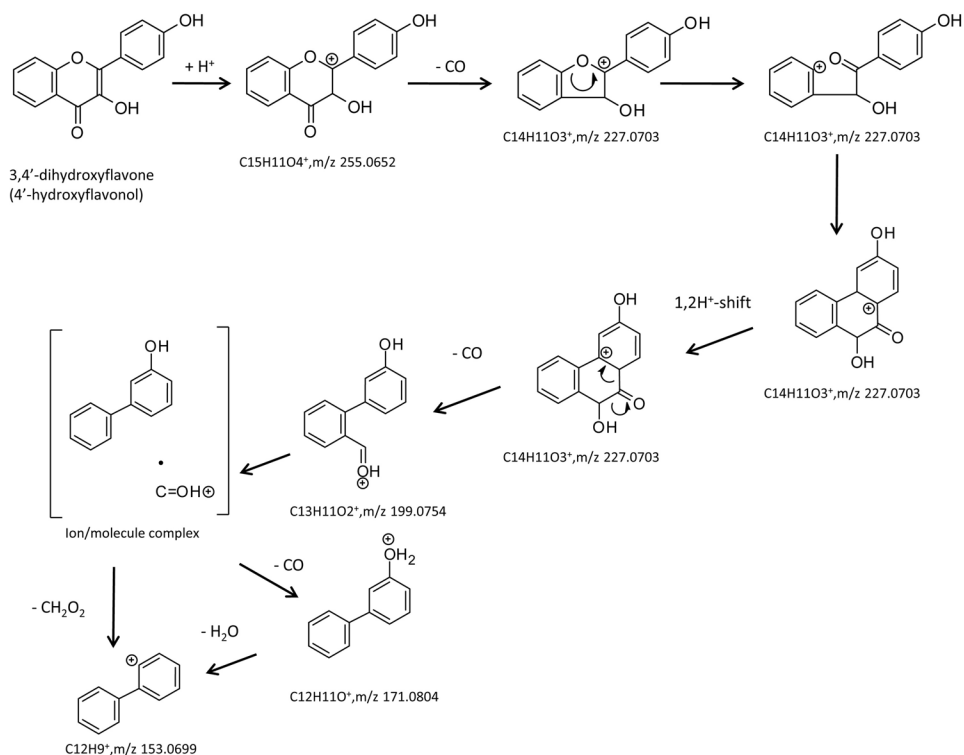


**Supplemental Figure 2:** MS<sup>n</sup> patterns in time and at different concentrations and collision energies. MS<sup>2</sup> fragmentation spectra of 6,4'-dihydroxyflavone (1) are shown: A) in positive ionization mode, averaged 6 times (dark grey) or 10 times (light grey); MS<sup>2</sup> spectra of 4'-hydroxyflavonol (6) are displayed: B) in negative ionization mode at 0 months (dark grey), 3 months (light grey), and 5 months (grey); C) in positive ionization mode at 0 months (dark grey), 3 months (light grey), and 5 months (grey); D) in negative ionization mode, at a normalized collision energy of 30% (dark grey), 35% (light grey), and 40% (grey). MS<sup>2</sup> and MS<sup>3</sup> fragmentation spectra of quercetin-3-O-glucoside (29) are shown: E) MS<sup>2</sup> spectra in negative ionization mode at a normalized collision energy of 30% (dark grey), 35% (light grey), and 40% (grey); F) MS<sup>3</sup> spectra of the C15H9O7 fragment ion in negative ionization mode at a normalized collision energy of 30% (dark grey), 35% (light grey), and 40% (grey). The average of six spectra is shown and the error bars represent a 95% confidence interval.



**Supplemental Figure 3:** Fragmentation spectra of the two positional isomers quercetin (8) (dark grey) and morin (7) (light grey) differing in the location of one hydroxyl group,  $MS^2$  in both negative (A) and positive (B) ionization mode. The average of six spectra is shown and the error bars represent a 95% confidence interval.





**Supplemental Figure 4:** Proposed fragmentation pathway leading to the C12H9 fragment ion for 3,4'-dihydroxyflavone in positive ionization mode.

**Table S-1:** The complete list of polyphenolic compounds that were fragmented in positive and negative ionization mode was collected in this Excel file. The file is available at <http://library.wur.nl/WebQuery/edepot/216854> and a direct link is: <http://edepot.wur.nl/216856>

**Table S-2:** The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 dihydroxylated flavones (1-6 in Fig. 1 of article) collected in negative ionization mode. The isomers share the same parent ion ([C<sub>15</sub>H<sub>9</sub>O<sub>4</sub>]<sup>-</sup>, m/z of 253.0506) that was selected for fragmentation in MS<sup>2</sup>. The flavone backbone is substituted with hydroxyl groups on the rings and sites specified in the table. The \* marks radical fragments.

1	2	3	4	5	6	Compound #
A,B	A,B	A	B,C	A,C	B,C	Rings/C's modified
6,4'	6,2'	6,7	3,2'	3,7	3,4'	Loss
				C15H8O3		[M-H-O]-
	C15H7O3					[M-H-OH]-
	C15H5O3					[M-H-OH3]-
C14H9O3	C14H9O3	C14H9O3		C14H9O3	C14H9O3	[M-H-CO]-
C14H9O2	C14H9O2	C14H9O2	C14H9O2	C14H9O2	C14H9O2	[M-H-CO2]-
C14H8O3*		C14H8O3*		C14H8O3*		[M-H-CHO*]-
C14H8O2*	C14H8O2*	C14H8O2*	C14H8O2*	C14H8O2*	C14H8O2*	[M-H-CHO2*]-
C14H7O3		C14H7O3		C14H7O3		[M-H-CH2O]-
C14H7O2	C14H7O2	C14H7O2		C14H7O2		[M-H-CH2O2]-
		C14H6O3*				[M-H-CH3O*]-
C13H9O2	C13H9O2	C13H9O2	C13H9O2	C13H9O2	C13H9O2	[M-H-2CO]-
C13H9O	C13H9O	C13H9O	C13H9O	C13H9O	C13H9O	[M-H-CO2-CO]-
	C13H9	C13H9	C13H9	C13H9		[M-H-2CO2]-
C13H8O2*		C13H8O2*		C13H8O2*		[M-H-CHO*-CO]-
C13H8O*	C13H8O*	C13H8O*		C13H8O*	C13H8O*	[M-H-CHO*-CO2]-
	C13H7O3					[M-H-C2H2O]-
C13H7O2		C13H7O2	C13H7O2	C13H7O2	C13H7O2	[M-H-2CHO]-
C13H7O	C13H7O	C13H7O	C13H7O	C13H7O	C13H7O	[M-H-CHO2-CO2]-
C13H6O3*				C13H6O3*		[M-H-C2H3O*]-
C12H9O	C12H9O	C12H9O	C12H9O	C12H9O	C12H9O	[M-H-3CO]-
C12H9	C12H9	C12H9	C12H9	C12H9	C12H9	[M-H-CO2-2CO]-
					C12H7O2	[M-H-C3H2O2]-
C12H7O	C12H7O				C12H7O	[M-H-CH2O2-CO]-
	C12H7		C12H7	C12H7	C12H7	[M-H-CH2O2-CO2]-
				C12H6O2*		[M-H-CHO*-C2H2O]-
		C11H9O				[M-H-C4O3]-
		C11H9			C11H9	[M-H-4CO]-
				C10H5O		RDA
	C9H4O3*			C9H4O3*		[M-H-B ring*]
			C8H5O2		C8H5O2	RDA 1,3B-
	C8H5O			C8H5O		RDA 1,3B-
C8H4O2*						RDA
C7H5O2					C7H5O2	RDA 0,2B-
						RDA
				C7H4O2*		RDA
	C7H3O3			C7H3O3		RDA
C6H5O3						RDA
	C6H5O		C6H5O	C6H5O		RDA
				C6H4O*	C6H4O*	RDA
	C6H4O2*					RDA
C6H3O2	C6H3O2					RDA
C5H3O	C5H3O			C5H3O		RDA
C4H5O						RDA

**Table S-3:** The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 dihydroxylated flavones (1-6 in Fig. 1 of article) collected in positive ionization mode. The isomers share the same parent ion formula ( $[C_{15}H_{11}O_4]^+$ ,  $m/z$  of 255.0652) that was selected for fragmentation in MS<sup>2</sup>. The flavone backbone is substituted with hydroxyl groups on the rings and sites specified in the table. The \* marks radical fragments.

1	2	3	4	5	6	Compound #
A,B	A,B	A	B,C	A,C	B,C	Rings/C's modified
6,2'	6,4'	6,7	3,2'	3,7	3,4'	Loss
		C15H9O3	C15H9O3			[M+H-H2O]+
C14H11O3	C14H11O3	C14H11O3			C14H11O3	[M+H-CO]+
				C14H10O3*	C14H10O3*	[M+H-CHO]+
		C14H9O2	C14H9O2	C14H9O2	C14H9O2	[M+H-CO-H2O]+
					C14H11O	[M+H-CO3]+
		C14H9O	C14H9O		C14H9O	[M+H-CO2-H2O]+
C13H11O2	C13H11O2	C13H11O2		C13H11O2	C13H11O2	[M+H-2CO]+
	C13H9O	C13H9O	C13H9O	C13H9O	C13H9O	[M+H-2CO-H2O]+
C13H9	C13H9	C13H9	C13H9	C13H9	C13H9	[M+H-CO-CO2-H2O]+
	C12H11O		C12H11O	C12H11O	C12H11O	[M+H-3CO]+
			C12H11			[M+H-CO2-2CO]+
C12H10O3*						[M+H-C3HO*]+
		C12H9O2				[M+H-CO-C2H2O]+
		C12H9O	C12H9O			[M+H-CO2-C2H2O]+
C12H9	C12H9	C12H9	C12H9	C12H9	C12H9	[M+H-3CO-H2O]+
			C11H11		C11H11	[M+H-4CO]+
				C11H9O		[M+H-2CO-C2H2O]+
		C11H9				[M+H-CO-CO2-C2H2O]+
			C10H8*			[M+H-C5H3O4*]
			C9H7O		C9H7O	[M+H-C6H4O3]
			C9H7			[M+H-C6H4O4]
C8H7O	C8H7O					RDA 1,3B+
				C8H5O2		RDA 0,3B+
			C8H5O2		C8H5O2	RDA 0,2A+-2H and/or 1,3B+-2H
C7H7						RDA -CO
		C7H5O4				RDA 1,3A+
C7H5O3	C7H5O3			C7H5O3		RDA 1,3A+
				C7H5O2		RDA 0,3A+
C7H5O2						RDA
			C7H5O2		C7H5O2	RDA 1,3A+
		C7H5O	C7H5O	C7H5O	C7H5O	RDA
		C7H3O3				RDA
C6H5O2						RDA 1,4A+
					C6H5O	RDA 1,4A+
		C6H3O2				RDA-CO
			C6H5	C6H5		RDA-CO
C5H5O						RDA 1,4A+-CO
		C5H3O				RDA -2CO

**Table S-4:** The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 6 monomethoxylated flavones (9-14 in Fig. 1 of article) collected in positive ionization mode. The isomers share the same parent ion formula ([C<sub>16</sub>H<sub>13</sub>O<sub>3</sub>]<sup>+</sup>, m/z of 253.0859) that was selected for fragmentation in MS<sup>2</sup>. The flavone backbone is substituted with methoxyl groups on the rings and sites specified in the table. The \* marks radical fragments.

9	10	11	Compound #
C	A	A	Rings/C modified
3	5	6	Loss
C15H10O3*	C15H10O3*	C15H10O3*	[M+H-CH3*] <sup>+</sup>
C15H9O3			[M+H-CH4] <sup>+</sup>
			[M+H-CH4O] <sup>+</sup>
		C14H13O	[M+H-2CO] <sup>+</sup>
C14H10O2*	C14H10O2*	C14H10O2*	[M+H-CH3*-CO] <sup>+</sup>
C14H9O2			[M+H-CH3*-CHO*] <sup>+</sup>
			[M+H-CH3*-CHO2*] <sup>+</sup>
		C14H8O*	[M+H-CH3*-CH2O2] <sup>+</sup>
			[M+H-CH3*-CO3] <sup>+</sup>
	C13H10O*	C13H10O*	[M+H-CH3*-2CO] <sup>+</sup>
C13H9O	C13H9O	C13H9O	[M+H-CH3*-CO*-CO] <sup>+</sup>
		C13H13	[M+H-CO-CO2] <sup>+</sup>
			[M+H-CH3*-CO-CO2] <sup>+</sup>
C13H9			[M+H-CH3*-CHO*-CO2] <sup>+</sup>
		C13H8*	[M+H-CH3*-CHO-CHO2] <sup>+</sup>
			[M+H-CH3*-C3HO2*] <sup>+</sup>
	C12H10*	C12H10*	[M+H-CH3*-3CO] <sup>+</sup>
C12H9	C12H9	C12H9	[M+H-CH3-2CO-CHO] <sup>+</sup>
			[M+H-C5H3O3] <sup>+</sup>
C9H7O2			[M+H-C7H6O] <sup>+</sup>
			[M+H-C7H4O2] <sup>+</sup>
		C9H5O	[M+H-C7H8O2] <sup>+</sup>
		C8H7O3	RDA 1,3B <sup>+</sup>
C8H6O*			RDA 1,3B <sup>+</sup> *
	C7H4O3*		RDA 1,3B <sup>+</sup> -CH3*
	C7H7O2	C7H7O2	RDA
			RDA 1,3B <sup>+</sup> -CO
			RDA 1,3B <sup>+</sup>
			RDA 1,3B <sup>+</sup> *
	C7H4O*		RDA [M+H-C8H9O2]
		C6H7O	RDA -CO
	C6H4O2*	C6H4O2*	RDA 1,3B <sup>+</sup> -CH3*
			RDA 1,3B <sup>+</sup> *-CO

12	13	14	Compound #
A	B	B	Rings/C modified
7	2'	4'	Loss
C15H10O3*	C15H10O3*	C15H10O3*	[M+H-CH3*] <sup>+</sup>
	C15H9O3		[M+H-CH4] <sup>+</sup>
	C15H9O2		[M+H-CH4O] <sup>+</sup>

C14H13O			[M+H-2CO] <sup>+</sup>
C14H10O2*	C14H10O2*	C14H10O2*	[M+H-CH3*-CO] <sup>+</sup>
			[M+H-CH3*-CHO*] <sup>+</sup>
	C14H9O		[M+H-CH3*-CHO2*] <sup>+</sup>
C14H8O*			[M+H-CH3*-CH2O2] <sup>+</sup>
		C14H10*	[M+H-CH3*-CO3] <sup>+</sup>
C13H10O*			[M+H-CH3*-2CO] <sup>+</sup>
C13H9O	C13H9O	C13H9O	[M+H-CH3*-CO*-CO] <sup>+</sup>
			[M+H-CO-CO2] <sup>+</sup>
		C13H10*	[M+H-CH3*-CO-CO2] <sup>+</sup>
	C13H9	C13H9	[M+H-CH3*-CHO*-CO2] <sup>+</sup>
C13H8*			[M+H-CH3*-CHO-CHO2] <sup>+</sup>
		C12H9O	[M+H-CH3*-C3HO2*] <sup>+</sup>
C12H10*			[M+H-CH3*-3CO] <sup>+</sup>
C12H9	C12H9	C12H9	[M+H-CH3-2CO-CHO] <sup>+</sup>
	C11H10	C11H10	[M+H-C5H3O3] <sup>+</sup>
	C9H7O2		[M+H-C7H6O] <sup>+</sup>
		C9H9O	[M+H-C7H4O2] <sup>+</sup>
C9H5O			[M+H-C7H8O2] <sup>+</sup>
C8H7O3			RDA 1,3B <sup>+</sup>
	C8H6O*		RDA 1,3B+*
			RDA 1,3B+-CH3*
C7H7O2			RDA
C7H7O			RDA 1,3B+-CO
	C7H5O2	C7H5O2	RDA 1,3B <sup>+</sup>
	C7H4O2*		RDA 1,3B+*
			RDA [M+H-C8H9O2]
C6H7O			RDA -CO
C6H4O2*			RDA 1,3B+-CH3*
	C6H4O*		RDA 1,3B+*-CO

**Table S-5 (starting on next page):** The fragments in the MS<sup>2</sup>-MS<sup>5</sup> spectra of 7 dimethoxylated flavones (15-21 in Fig. 1 of article) collected in positive ionization mode. The isomers share the same parent ion formula ( $[C_{17}H_{15}O_4]^+$ ,  $m/z$  of 283.0965) that was selected for fragmentation in MS<sup>2</sup>. The flavone backbone is substituted with methoxyl groups on the rings and sites specified in the table. The \* marks radical fragments.

15	16	17	18	Compound #
A,C	A,C	C	C	Rings/C's modified
3,6	3,7	5,7	7,8	Loss
C16H12O4*	C16H12O4*	C16H12O4*	C16H12O4*	[M+H-CH3*]+
C16H11O4	C16H11O4			[M+H-CH4]+
		C16H11O3		[M+H-CH4O]+
				[M+H-CH5O*]+
				[M+H-CH6O]+
C15H8O4*	C15H8O4*			[M+H-CH3*-CH4]+
C15H12O3*	C15H12O3*			[M+H-CH3*-CO]+
C15H11O3		C15H11O3	C15H11O3	[M+H-CH3*-CHO*]+
		C15H11O2		[M+H-CH3*-CHO2*]+
C15H10O3*		C15H10O3*		[M+H-CH3*-CH2O]+
				[M+H-CH3*-CHO3*]+
			C15H10O2*	[M+H-CH3*-CH2O2]+
				[M+H-CH3*-CH3*]+
				[M+H-CH3*-CH3O*]+
			C15H9O2	[M+H-2CH3O]+
			C15H9O	[M+H-CH3*-CH3O3*]+
				[M+H-CH3*-CH4]+
				[M+H-CH3*-CH4O]+
			C15H8O*	[M+H-CH3*-CH4O3]+
		C14H11O2		[M+H-CH3*-CHO*-CO]+
		C14H10O2*		[M+H-CH3*-2CHO]+
C14H9O3	C14H9O3	C14H9O3		[M+H-2CH3***-CO]+
C14H9O2	C14H9O2			[M+H-2CH3***-CO2]+
				[M+H-2CH3***-CO3]+
C14H8O3	C14H8O3	C14H8O3		[M+H-2CH3***-CHO]+
C14H11O2	C14H11O2			[M+H-CH3*-CHO2*]+
		C14H11O		[M+H-CH3*-CHO3*]+
				[M+H-CH3*-CHO4*]+
			C14H10O*	[M+H-CH3*-CH2O2-CO]+
				[M+H-CH3*-CH4-CO]+
				[M+H-CH3*-CH4O-CO]+
C13H11O	C13H11O	C13H11O	C13H11O	[M+H-CH3*-CHO2*-CO]+
				[M+H-CH3*-CHO4*-CO]+
		C13H10O*		[M+H-CH3*-CHO-2CO]+
			C13H10*	[M+H-CH3*-CH2O2-2CO]+
C13H9O2	C13H9O2	C13H9O2	C13H9O2	[M+H-2CH3***-2CO]+
C13H9O	C13H9O	C13H9O		[M+H-2CH3***-CO-CO2]+
C13H9		C13H9	C13H9	[M+H-2CH3***-2CO2]+
C13H8O2*	C13H8O2*			[M+H-2CH3*-CHO-CO]
				[M+H-CH3*-CH4O-2CO]+
C13H7O2	C13H7O2			[M+H-2CH3*-2CHO*]+
				[M+H-CH3*-3CO]+
	C12H9O	C12H9O		[M+H-2CH3***-3CO]+
C12H8O*	C12H8O*			[M+H-2CH3-CHO-2CO]+
		C12H11O		[M+H-CH3*-CO-2CHO*]+
		C12H11		[M+H-CH3*-CHO*-3CO]+

		C12H10*		[M+H-CH3*-2CHO-2CO]+
C12H9	C12H9	C12H9		[M+H-2CH3** <sub>2</sub> -2CO-CO <sub>2</sub> ]+
				[M+H-CH3*-CH4O-3CO]+
		C11H13		[M+H-C6H4O4]+
	C11H9	C11H9		[M+H-2CH3-4CO]+
C11H8*	C11H8*			[M+H-2CH3-CHO-3CO]+
	C11H7		C11H7	[M+H-2CH3-2CHO-2CO]+
				[M+H-2CH3*-C5H3O]+
				RDA <sup>1,3</sup> B <sup>+</sup> -CH3
				RDA
				RDA <sup>1,3</sup> B <sup>+</sup> -CH3*
	C9H7O2			RDA
				RDA
				RDA
		C9H5O		RDA
				RDA
				RDA
C8H7O2	C8H7O2			RDA
				RDA
				RDA
				RDA
			C8H5O4	RDA <sup>1,3</sup> A <sup>++</sup> -CH3*
		C8H5O3		RDA
				RDA
		C8H7		RDA
				RDA
		C7H7O4		RDA
				RDA <sup>1,3</sup> A <sup>+</sup> +2H
				RDA
				RDA
				RDA
		C7H5O3	C7H5O3	RDA C8H5O4-CO
				RDA
C7H5O	C7H5O		C7H5O	RDA
			C7H4O2*	RDA
			C7H3O2	RDA
		C6H5O2	C6H5O2	RDA C8H5O4-2CO
		C6H5O		RDA C7H5O3-CO <sub>2</sub>
		C6H4O2*	C6H4O2*	RDA
		C6H3O3		RDA
C6H3O2			C6H3O2	RDA C7H6O2-CH3*
			C6H4O*	RDA C7H4O2*-CO
		C5H5O	C5H5O	RDA C7H5O3-CO <sub>2</sub>
				RDA C6H5O-CO
		C3HO2		RDA

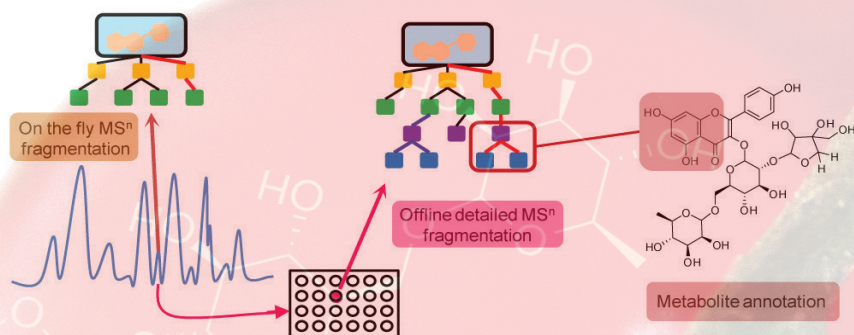
19	20	21	Compound #
B,C	B,C	A,B	Rings/C's modified
3,2'	3,4'	5,2'	Loss
C16H12O4*	C16H12O4*	C16H12O4*	[M+H-CH3*]+
	C16H11O4		[M+H-CH4]+
C16H11O3		C16H11O3	[M+H-CH4O]+
C16H10O3*		C16H10O3*	[M+H-CH5O*]+
		C16H9O3	[M+H-CH6O]+
			[M+H-CH3*-CH4]+
C15H12O3*	C15H12O3*	C15H12O3*	[M+H-CH3*-CO]+
C15H11O3		C15H11O3	[M+H-CH3*-CHO*]+
C15H11O2		C15H11O2	[M+H-CH3*-CHO2*]+
			[M+H-CH3*-CH2O]+
C15H11O			[M+H-CH3*-CHO3*]+
C15H10O2*		C15H10O2*	[M+H-CH3*-CH2O2]+
C15H9O4	C15H9O4		[M+H-CH3*-CH3*]+
C15H9O3	C15H9O3		[M+H-CH3*-CH3O*]+
C15H9O2		C15H9O2	[M+H-2CH3O]+
C15H9O		C15H9O	[M+H-CH3*-CH3O3*]+
C15H8O4*	C15H8O4*	C15H8O4*	[M+H-CH3*-CH4]+
C15H8O3*		C15H8O3*	[M+H-CH3*-CH4O]+
			[M+H-CH3*-CH4O3]+
			[M+H-CH3*-CHO*-CO]+
			[M+H-CH3*-2CHO]+
C14H9O3	C14H9O3	C14H9O3	[M+H-2CH3***-CO]+
C14H9O2	C14H9O2		[M+H-2CH3***-CO2]+
C14H9O		C14H9O	[M+H-2CH3***-CO3]+
		C14H8O3	[M+H-2CH3***-CHO]+
C14H11O2			[M+H-CH3*-CHO2*]+
C14H11O		C14H11O	[M+H-CH3*-CHO3*]+
C14H11			[M+H-CH3*-CHO4*]+
C14H10O*		C14H10O*	[M+H-CH3*-CH2O2-CO]+
	C14H8O3*		[M+H-CH3*-CH4-CO]+
C14H8O2*		C14H8O2*	[M+H-CH3*-CH4O-CO]+
C13H11O			[M+H-CH3*-CHO2*-CO]+
C13H11		C13H11	[M+H-CH3*-CHO4*-CO]+
			[M+H-CH3*-CHO-2CO]+
		C13H10*	[M+H-CH3*-CH2O2-2CO]+
C13H9O2	C13H9O2		[M+H-2CH3***-2CO]+
C13H9O	C13H9O		[M+H-2CH3***-CO-CO2]+
C13H9		C13H9	[M+H-2CH3***-2CO2]+
	C13H8O2*		[M+H-2CH3*-CHO-CO]
C13H8O*		C13H8O*	[M+H-CH3*-CH4O-2CO]+
	C13H7O2		[M+H-2CH3*-2CHO*]+
	C12H12O*		[M+H-CH3*-3CO]+
	C12H9O		[M+H-2CH3***-3CO]+
	C12H8O*		[M+H-2CH3-CHO-2CO]+
			[M+H-CH3*-CO-2CHO*]+
			[M+H-CH3*-CHO*-3CO]+



			[M+H-CH <sub>3</sub> *-2CHO-2CO]+
C12H9			[M+H-2CH <sub>3</sub> **_2CO-CO <sub>2</sub> ]+
C12H8*		C12H8*	[M+H-CH <sub>3</sub> *-CH <sub>4</sub> O-3CO]+
			[M+H-C <sub>6</sub> H <sub>4</sub> O <sub>4</sub> ]+
	C11H9		[M+H-2CH <sub>3</sub> -4CO]+
	C11H8*		[M+H-2CH <sub>3</sub> -CHO-3CO]+
			[M+H-2CH <sub>3</sub> -2CHO-2CO]+
C10H6O <sub>3</sub> *			[M+H-2CH <sub>3</sub> *-C <sub>5</sub> H <sub>3</sub> O]+
	C9H9O		RDA <sup>1,3</sup> B <sup>+</sup> -CH <sub>3</sub>
C9H8O <sub>2</sub> *			RDA
	C9H8O*		RDA <sup>1,3</sup> B <sup>+</sup> -CH <sub>3</sub> *
			RDA
C9H6O <sub>2</sub> *			RDA
C9H5O <sub>2</sub>			RDA
			RDA
C8H8O*			RDA
	C8H7O <sub>3</sub>		RDA
			RDA
C8H7O			RDA
		C8H6O <sub>3</sub> *	RDA
C8H6O*			RDA
			RDA <sup>1,3</sup> A <sup>++</sup> -CH <sub>3</sub> *
			RDA
C8H5O <sub>2</sub>			RDA
			RDA
C7H9			RDA
			RDA
	C7H7O <sub>2</sub>		RDA <sup>1,3</sup> A <sup>++</sup> +2H
C7H8*			RDA
C7H7			RDA
		C7H6O <sub>2</sub> *	RDA
C7H6*			RDA
		C7H5O <sub>3</sub>	RDA C <sub>8</sub> H <sub>5</sub> O <sub>4</sub> -CO
C7H5O <sub>2</sub>			RDA
C7H5O			RDA
			RDA
			RDA
			RDA C <sub>8</sub> H <sub>5</sub> O <sub>4</sub> -2CO
C6H5O		C6H5O	RDA C <sub>7</sub> H <sub>5</sub> O <sub>3</sub> -CO <sub>2</sub>
	C6H4O <sub>2</sub> *	C6H4O <sub>2</sub> *	RDA
		C6H3O <sub>3</sub>	RDA
		C6H3O <sub>2</sub>	RDA C <sub>7</sub> H <sub>6</sub> O <sub>2</sub> -CH <sub>3</sub> *
			RDA C <sub>7</sub> H <sub>4</sub> O <sub>2</sub> *-CO
		C5H5O	RDA C <sub>7</sub> H <sub>5</sub> O <sub>3</sub> -CO <sub>2</sub>
C5H5			RDA C <sub>6</sub> H <sub>5</sub> O-CO
			RDA



## Chapter 3: Spectral trees as a robust annotation tool in LC-MS based metabolomics



Justin J.J. van der Hooft, Jacques Vervoort, Raoul J. Bino,  
and Ric C.H. de Vos

This chapter was published as research article in *Metabolomics*, 2012, Volume 8 (4), pp 691–703, DOI: 10.1007/s11306-011-0363-7

### *Abstract*

The identification of large series of metabolites detectable by mass spectrometry (MS) in crude extracts is a challenging task. In order to test and apply the so-called multistage mass spectrometry (MS<sup>n</sup>) spectral tree approach as tool in metabolite identification in complex sample extracts, we firstly performed liquid chromatography (LC) with online electrospray ionization (ESI)-MS<sup>n</sup>, with the use of crude extracts from both tomato fruit and *Arabidopsis* leaf. Secondly, the extracts were automatically fractionated by a NanoMate LC-fraction collector/injection robot (Advion) and selected LC-fractions were subsequently analyzed with the use of nanospray-direct infusion to generate offline in-depth MS<sup>n</sup> spectral trees at high mass resolution. Characterization and subsequent annotation of metabolites was achieved by detailed analysis of the MS<sup>n</sup> spectral trees, thereby focusing on two major plant secondary metabolite classes: phenolics and glucosinolates. Following this approach, we were able to discriminate all selected flavonoid glycosides, based on their unique MS<sup>n</sup> fragmentation patterns in either negative or positive ionization mode. As a proof of principle, we report here 127 annotated metabolites in the tomato and *Arabidopsis* extracts, including 21 novel metabolites. Our results indicate that online LC-MS<sup>n</sup> fragmentation in combination with databases of in-depth spectral trees generated offline can provide a fast and reliable characterization and annotation of metabolites present in complex crude extracts such as those from plants.

Key words: annotation, fragmentation, LC-MS, metabolomics, metabolite identification, structural elucidation

## Introduction

The characterization, annotation and structural elucidation of metabolites in crude extracts are of great importance in MS based metabolomics.<sup>1-5</sup> The enormous chemical diversity and the vast amount of structurally related and isomeric compounds make the precise identification of metabolites a difficult task. Crude extracts from plants are particularly known to contain a wide array of metabolites, specifically those involved in secondary metabolism, like phenylpropanoids, glucosinolates, alkaloids, terpenoids, and the economically important compound class of flavonoids. Comprehensive, untargeted LC-MS profiling of such plant extracts, including those from well studied species like tomato (*Lycopersicon esculentum*) and *Arabidopsis thaliana*, typically results in compounds lists in which more than half of the detected metabolites are still completely unknown (MSI metabolite identification level 4) or at most only partially characterized (MSI metabolite identification level 3).<sup>6-9</sup>

Tomato fruit peel has been subjected to metabolomics studies frequently.<sup>8, 10-12</sup> Likewise, leaves from *A. thaliana* plants have been frequently investigated with the use of LC-MS.<sup>6, 13-15</sup> Nevertheless, far from all metabolites have been unambiguously elucidated and thus completely annotated even in these important model species.<sup>16</sup> In recent studies, the role of LC-MS and especially the need for MS fragmentation in the metabolite identification process is highlighted.<sup>17, 18</sup> For example, MS<sup>2</sup> fragmentation with a time of flight (TOF)-MS and an Ion trap MS revealed 21 novel compounds in fruits of tomato.<sup>10</sup> Another study using a Triple Quad (MS<sup>2</sup>) and an Orbitrap system (up to MS<sup>3</sup>) yielded three new compounds in tomato.<sup>12</sup> In addition, Ion trap MS fragmentation up to MS<sup>4</sup> was used to detect 13 novel tomato seed compounds.<sup>19</sup> Moreover, recently a method was described,<sup>20</sup> based on Ion trap MS fragmentation up to MS<sup>3</sup>, which enabled the simultaneous detection of more than 30 phenolic compounds and 12 glucosinolates in *Brassica rapa* in a quantitative manner. These examples underline the power of MS fragmentation in metabolite annotation.

Since its introduction, the Ion trap - Orbitrap platform has found many applications in MS based metabolomics<sup>21, 22</sup> and proteomics,<sup>23</sup> using the MS<sup>n</sup>

fragmentation abilities of the Ion trap in combination with the accurate mass provided by the Orbitrap FTMS. The obtained accurate mass enables rapid assignment of elemental formulas to detected peptides or metabolites and their fragments present in biological extracts.

On-the-fly fragmentation experiments during LC-MS are hampered by either a limited fragmentation depth, which is usually  $MS^2$  and at maximum  $MS^3$ , or a nominal mass resolution of  $MS^n$  fragments, the latter complicating elemental formula calculation of the parent molecule and its fragments. As a result, many metabolites detected in LC-MS profiles are only partially annotated and isomeric compounds are rarely differentiated, leading to a vast number of tentatively identified metabolites present in literature. Detailed metabolite comparison of complex extracts and biological interpretation of differential profiles, however, is only possible with a more precise annotation of metabolites.

We recently described a method for highly reproducible and in-depth accurate mass  $MS^n$  fragmentation<sup>24</sup> of metabolites resulting in so-called spectral trees.<sup>25</sup> With the use of nanospray infusion into a LTQ - Orbitrap hybrid mass spectrometer, we could discriminate 119 out of 121 tested polyphenolic compounds including different series of isomers.<sup>24</sup> We concluded that the  $MS^n$  spectra can be used for the differentiation and identification of metabolite structures, as they can provide not only unique fragments, but also specific differences in the relative intensities of the fragment ions.

The aim of the present study was to test the applicability of accurate mass  $MS^n$  spectral trees as a tool in the identification and structural elucidation of metabolites detected during LC-MS profiling of complex sample matrices, such as crude plant extracts. As proof of principle, we focused on phenolic compounds, in order to enable comparison of fragmentation spectra between standards<sup>24</sup> and metabolites present in crude extracts. Since extracts of most plant species contain a variety of secondary metabolites that usually exist in different isomeric forms, such as conjugated flavonoids, the correct annotation of these isobaric secondary metabolites represents a technological challenge.

Here, we tested the ability of LC- $MS^n$  spectral trees to identify and discriminate

flavonoid species in tomato and *Arabidopsis thaliana*, two plant species most frequently used in metabolomics and genomics studies. We firstly performed a detailed MS<sup>n</sup> analysis of phenolic compounds in tomato fruit, and then used *Arabidopsis* leaf as an example to quickly characterize and annotate a number of phenolic compounds as well as glucosinolates. Firstly, the combination of the MS<sup>n</sup> fragmentation ability of the Ion trap MS and the high mass accuracy of the Orbitrap FTMS was used to generate (partial) spectral tree data online during LC-MS analysis in an unbiased manner. Secondly, with the use of a NanoMate fraction collector/injection robot coupled between the LC column and the ionization source, we also generated offline data-directed in-depth MS<sup>n</sup> spectra of selected chromatographic peaks. Following these online and offline MS<sup>n</sup> approaches, we set up a generic procedure that can comprehensively retrieve structural information of a large range of metabolites present in crude plant extracts in an informative and robust manner.

## *Material and Methods*

### *Chemicals and plant material.*

*Chemicals:* Acetonitrile (HPLC grade) was obtained from Biosolve (Valkenswaard, The Netherlands), methanol (HPLC grade) from Merck-Schuchardt (Hohenbrunn, Germany), and formic acid (99-100) from VWR international S.A.S. (Briare, France). Ultrapure water was made in purification units present in-house.

*Plant harvesting and metabolite extraction:* Tomato and *Arabidopsis thaliana* material were grown in the greenhouse and directly after harvesting, the plant material was frozen, ground, and stored at -80°C. The frozen powder was then used to perform metabolite extractions, or firstly freeze-dried to obtain extra concentrated extracts. The extraction protocol was as described earlier<sup>26</sup> with slight modifications (Supplemental Text S1 in the Supporting Information).



*Analytical methods.*

*Online fragmentation using HPLC-PDA-ESI-MS<sup>n</sup>:* The set up consisted of an Accela HPLC tower connected to a LTQ/Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific). The LC conditions used were as described earlier<sup>8</sup> (details in Supplemental Text S1 in the Supporting Information). Eluting compounds were trapped within an LTQ Ion trap followed by automated MS<sup>n</sup> fragmentation. Subsequently, either the nominal or the accurate mass of the generated molecular fragments were recorded, using the Ion trap MS or the Orbitrap FTMS, respectively. MS settings and spectral tree topologies are given in Supplemental Text S1 in the Supporting Information.

*NanoMate LC-fractionation of plant extracts:* The HPLC-PDA-ESI-MS<sup>n</sup> system was adapted with a chip-based nano-electrospray ionization source / fractionation robot (NanoMate Triversa, Advion BioSciences) coupled between the PDA and the inlet of the Ion trap - Orbitrap hybrid instrument. Sample injection volume was 50  $\mu$ l. The gradient and flow conditions were the same as described above, with an additional 30  $\mu$ l/min 100% isopropanol added into the LC flow via a T-junction between the PDA and the NanoMate. The eluens flow was split by the NanoMate, at 219.5  $\mu$ L/min to the fraction collector and 0.5  $\mu$ L/min to the nano-electrospray source. LC-fractions were collected every 5 sec (i.e. 18  $\mu$ L) into a 384 wells plate (Twin tec, Eppendorf), cooled at 10 °C. After collection, 4  $\mu$ L of isopropanol was added to each well, in order to improve spray stability, and the plate was sealed. Details are given in Supplemental Text S1 in the Supporting Information.

*Offline MS<sup>n</sup> spectral tree generation:* The general analysis procedure and spectral tree topology have been described recently.<sup>24</sup> Shortly, the Ion trap was programmed to fragment a selected mass up to MS<sup>5</sup> level in a data dependent manner, thereby automatically selecting the three most intense ions in the MS<sup>n</sup> spectra for further fragmentation, after which the Orbitrap recorded the accurate mass of fragments generated. In the present study, the NanoMate robot was programmed to take up 8  $\mu$ L solvent from collected LC fractions, followed by direct infusion and nano-spraying into the MS (for details see Supplemental Text



1 in the Supporting Information).

*MS<sup>n</sup> data processing.*

Settings for automatic elemental formula calculation of accurate mass peaks obtained by Orbitrap experiments were: maximum mass accuracy deviation 10 ppm, charge state 1, maximum number of possible C, H, O, N, P, and S atoms were 80, 100, 50, 4, 2, and 4, respectively. The nitrogen rule was not applied, in order to include possible radical ions.<sup>24</sup>

With the use of the Xcalibur software, spectrum lists, including accurate *m/z* values, elemental formulas and relative intensities, were generated from the raw data files and exported into Excel. For online obtained spectral trees, the relative intensities detected in each individual scan were taken. For offline obtained spectral trees, the relative intensities of repetitive MS<sup>n</sup> spectra derived from the same (fragment) ion were averaged and plotted with a 95% confidence interval (i.e. two times standard deviation up and down). If the fragmented ion was still the base peak, i.e., the most intense peak within the spectrum, this mass peak was excluded. A threshold intensity of 3%, as compared to the base peak, was used for selecting masses in each fragmentation spectrum, like described before.<sup>24</sup> Elemental formulas of fragments were checked for correctness of the ring and double bond (RDB) factor, as well as for any violations with the parental formula. For the comparison of FTMS with Ion trap MS read-outs of fragmentation data, the accurate masses were converted into nominal masses.

MS<sup>n</sup> fragmentation patterns were considered identical when two criteria were met: I) all observed fragments (at the set threshold of >3% relative intensity) were present in both spectra; and in case of offline MS<sup>n</sup> all fragments should also be present in at least 2/3 of the repetitive scans, and II) the difference in relative intensities of all fragments present in the spectra was less than 20% (arithmetic difference); for instance, if the relative intensity of a specific fragment was 40% in spectrum 1 and 70% in spectrum 2, these two spectra were considered different.

*NMR identification of selected compounds.*

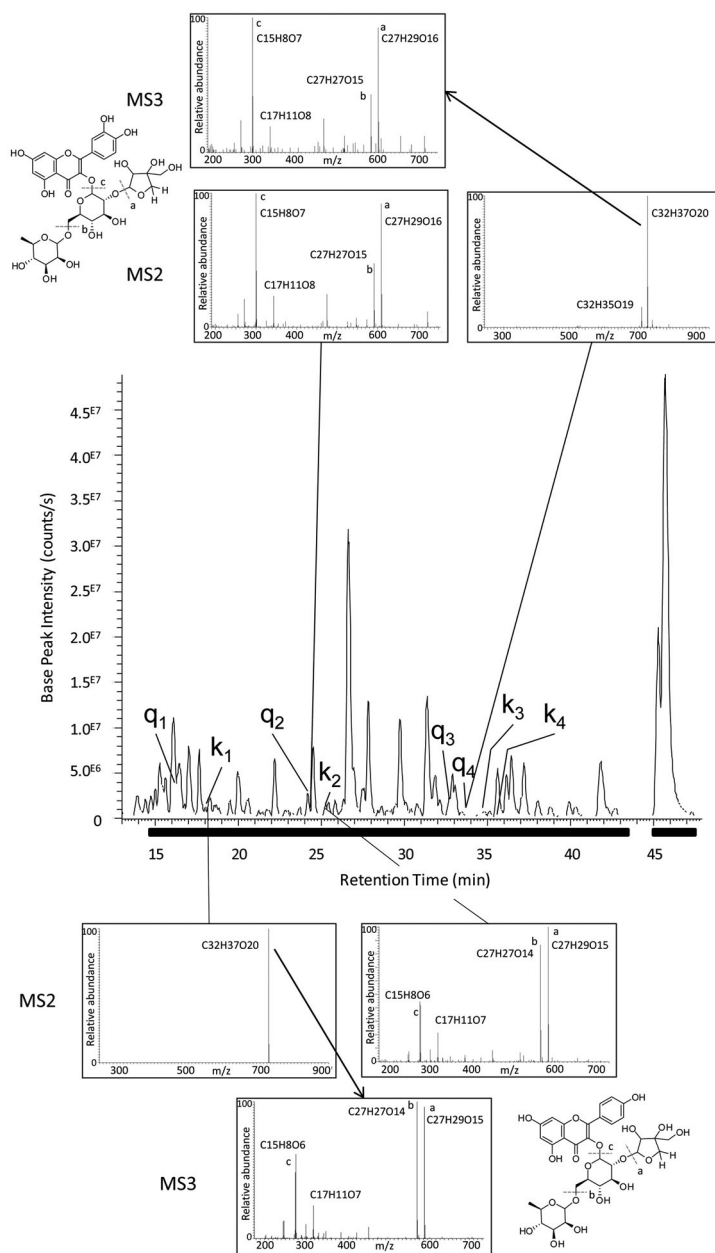
In order to verify compound annotations resulting from MS<sup>n</sup> experiments, compounds were manually collected by LC peak-fractionation of the crude plant extracts. The fractions were dried and redissolved in MeOD, followed by 1D-<sup>1</sup>H and 2D-COSY and 2D-TOCSY and 2D-HSQC NMR measurements on a 600 MHz NMR spectrometer (Bruker).<sup>4</sup>

*Results*

Metabolites in crude plant extracts were separated using reversed-phase C18 analytical chromatography, followed by untargeted online LC-MS<sup>n</sup> in negative electrospray ionization mode. Parts of the chromatograms were subsequently subjected to LC-fractionation using the NanoMate robot (Figure 1). Selected flavonoid glycosides, caffeoylquinic acids and, in case of *Arabidopsis*, glucosinolates (Figure 2), were fragmented offline in both negative and positive ionization mode, and the resulting spectral data was listed (Supplemental Table 1 in the Supporting Information). The on the fly LC-MS<sup>n</sup> fragmentation of metabolites is referred to as online LC-MS<sup>n</sup>, whereas the MS<sup>n</sup> fragmentation of LC-fractionated metabolites, using the NanoMate fractionation-injection robot, is referred to as offline MS<sup>n</sup> fragmentation. The mass accuracy obtained during both online LC-MS<sup>n</sup> and offline MS<sup>n</sup> analyses was always within 3 ppm, at all compound concentrations and MS<sup>n</sup> levels, and for full scan (MS1) always within 1.5 ppm, which is in agreement with our previous results.<sup>24</sup>

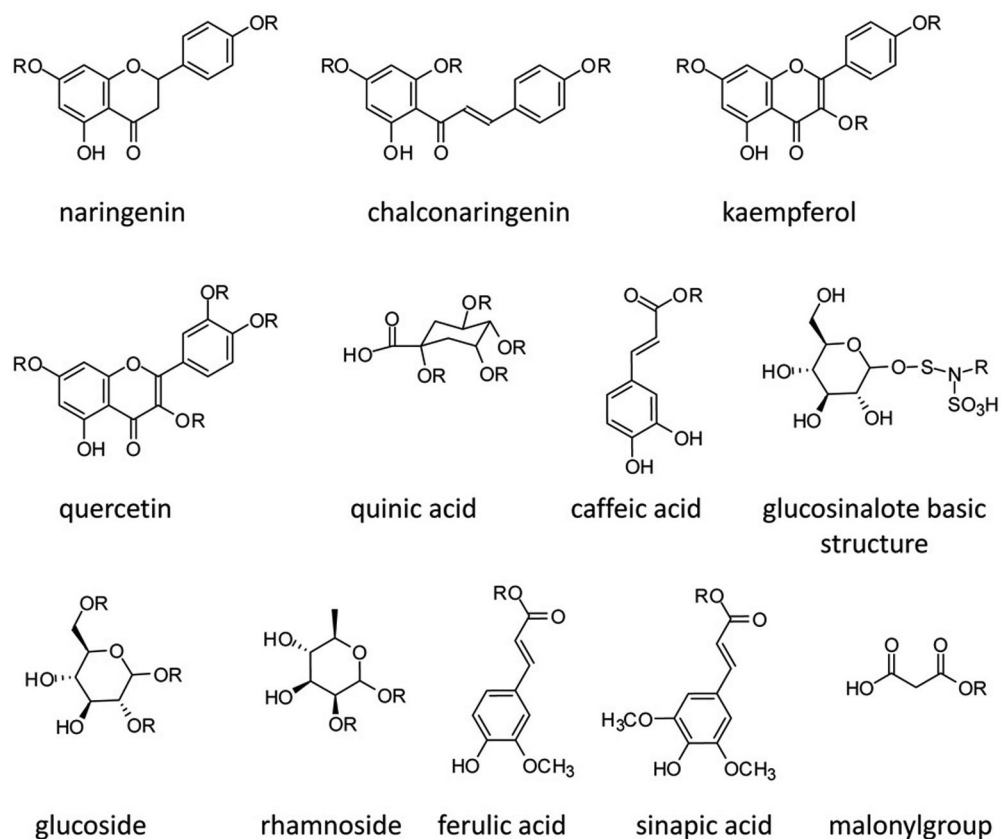
*Robustness and reproducibility of online and offline MS<sup>n</sup> spectra.*

In order to determine the similarity between high mass resolution MS<sup>n</sup> fragmentation patterns of metabolites and their fragments, the obtained MS<sup>n</sup> spectra were compared according to two criteria. These criteria were determined based on a large set of data acquired in our previous study using standard compounds<sup>24</sup> and data obtained within the present study (see Section on MS<sup>n</sup> data processing in Material and Methods). Multiple online LC-MS<sup>n</sup> spectra from the same compound can be obtained by continuously generating spectra from



**Figure 1:** Part of a tomato fruit LC-Orbitrap FTMS chromatogram in negative ionization mode. Black bars indicate the chromatographic sections fractionated each 5 seconds by the NanoMate, thereby filling up a 384 wells plate.  $q_1$ - $q_4$  and  $k_1$ - $k_4$  refer to four different quercetin- and four different kaempferol-glycosides, respectively. On-line MS<sup>2</sup> and MS<sup>3</sup> spectra of  $q_2$  and  $q_4$  and the structure of  $q_2$  (upper part) and on-line MS<sup>2</sup> and MS<sup>3</sup> spectra of  $k_1$  and  $k_2$  and the structure of  $k_2$  (lower part), are shown. The most likely fragmentation events causing the major fragments are indicated by dashed grey lines in the structures marked with corresponding letters.

its eluting chromatographic peak. This online spectra generation implies that metabolite spectra are obtained at different parent ion concentrations. Rutin (quercetin-3-*O*-rutinoside), a flavonoid-glycoside commonly occurring in plants, was taken as an example to test the reproducibility of online LC-MS<sup>n</sup> spectra.



**Figure 2:** Glucosinolate basic structure and core structures of several flavonoids and phenylpropanoids as well as their most common substituents are shown. The *R*'s indicate the most common sites for conjugations in plants.

During chromatography of a tomato fruit extract, four repetitive LC-MS<sup>3</sup> spectra of the quercetin fragment of rutin were obtained (Figure 3A). While the parent ion intensities ranged from 2E<sup>6</sup> - 4E<sup>7</sup>, the differences in relative intensities of all fragment ions present in the MS<sup>3</sup> spectra were always within 20% difference and in most cases differing less than 5%. The robustness of the online LC-MS<sup>n</sup> approach was further tested by comparing MS<sup>3</sup> spectra of the chromatographic peak of quercetin-3-*O*-(2''-*O*-apiofuranosyl-6''-*O*-rhamnosylglucoside), present in tomato fruit peel, obtained *in duplo* and with a 6 month time interval (Figure 3B). Even though the absolute intensities of the fragment ions varied slightly, the resulting four MS<sup>3</sup> spectra were all identical according to the two

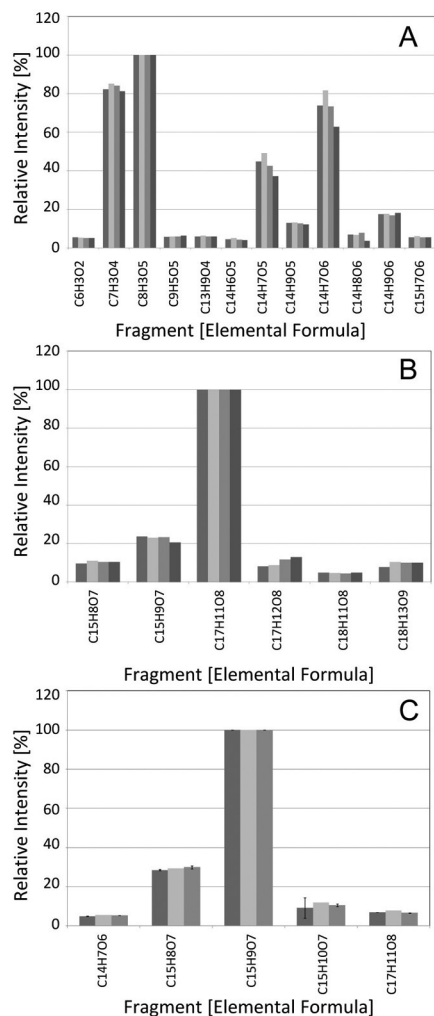
criteria. Subsequently, the spectra of rutin obtained by the online approach were compared with those of the offline MS<sup>n</sup> approach, as well as with those of its authentic standard measured by offline direct infusion<sup>24</sup> (Figure 3C). The fragmentation spectra were all identical. These results illustrate the robustness and reproducibility of the online high mass resolution LC-MS<sup>n</sup> approach when using the normalized fragmentation spectra for comparisons.

*Online LC-MS<sup>n</sup> and offline MS<sup>n</sup> spectral trees of crude plant extracts.*

Since the online high mass resolution LC-MS<sup>n</sup> approach proved to be robust, the method was tested for its applicability in the characterization and annotation of metabolites in crude plant extracts, using extracts from leaves of a mix of *Arabidopsis* ecotypes and from red tomato fruit peel. The *Arabidopsis* extract contained both phenolic compounds and glucosinolates, which were subjected to MS<sup>n</sup> fragmentation. In the tomato extract, we were able to detect and fragment a range of different flavonoid glycosides, mainly based on the aglycones quercetin, naringenin or chalconaringenin, and kaempferol. The fragmentation of these aglycone fragments was identical to the MS<sup>n</sup> spectral data from their respective reference compounds.<sup>24</sup> The resulting fragments of the analyzed metabolites from tomato and *Arabidopsis* are provided in Supplemental Table S1 in the Supporting Information.

The online LC-MS<sup>n</sup> approach resulted in 270 MS<sup>2</sup> scan events in the Orbitrap for a specific m/z value. As there are multiple series of isomers present in the extract, resulting in multiple scan events for some m/z values, the number of unique MS<sup>2</sup> scans is estimated to be around 450. In MS<sup>3</sup>, the number of scan events per unique m/z is 490, leading to an estimated total scan number of 600. The LC-MS<sup>n</sup> approach resulted in high-density robust fragmentation data of the crude extracts, which is underlined by the amount of accurate MS<sup>n</sup> spectra, and thus (partial) spectral trees, acquired within a one hour gradient.

The crude extracts were fractionated during HPLC separation into 5 seconds fractions using the NanoMate robot, and offline high resolution MS<sup>n</sup> spectral trees were subsequently generated for selected metabolites (Supplemental Table



**Figure 3:** Reproducibility of MS<sup>n</sup> spectra generated from tomato compounds. A) four successive online MS<sup>3</sup> spectra of the quercetin fragment, derived within the chromatographic peak of rutin, generated at retention time 26.64 min (start of peak, dark grey), 26.75 min (top of peak; light grey), 26.87 min (grey) and 26.99 min (end of peak, darkest grey); B) online MS<sup>3</sup> spectra of quercetin-3-O-(2''-apiofuranosyl-6''-rhamnosylglucoside) obtained at month 0 (dark grey and light grey; two repetitive LC-MS<sup>n</sup> runs) and month 6 (grey and darkest grey; two repetitive LC-MS<sup>n</sup> runs); C) MS<sup>2</sup> spectra of rutin obtained during offline MS<sup>n</sup> of collected LC-MS peak (dark grey, n=2), online during LC-MS<sup>n</sup> (light grey), and of the authentic standard (grey, n=3). The error bars represent two times the standard deviation.

S1 in the Supporting Information). Although most of the measured metabolite intensities ( $1E^5$ – $5E^6$  counts/s) were lower than those used for the reference compounds ( $1E^7$ – $1E^8$  counts/s),<sup>24</sup> detailed fragmentation spectra ranging from MS<sup>3</sup> up to MS<sup>5</sup>, depending on compound concentration and ionization efficiency of fragments, could be obtained for almost all flavonoids. In contrast, online accurate mass LC-MS fragmentation resulted in spectral trees of MS<sup>3</sup> at most, mainly due to time limitations during peak elution. Nevertheless, this online MS<sup>n</sup> fragmentation also generated relevant data of spectral similarities between

compounds. For example, of 4 quercetin (q<sub>1</sub>-q<sub>4</sub>) and 4 kaempferol (k<sub>1</sub>-k<sub>4</sub>) conjugates detected in tomato peel (Fig. 1), their 3 most complex glycosides (q<sub>1</sub>, q<sub>3</sub>, q<sub>4</sub>, and k<sub>1</sub>, k<sub>3</sub>, k<sub>4</sub>) produced MS<sup>3</sup> spectra that were similar to the MS<sup>2</sup> spectra of their lesser conjugated analogues (q<sub>2</sub> and k<sub>2</sub>, respectively, structures shown in Fig. 1). This result shows the ability of MS<sup>n</sup> to elucidate common substructures within related complex metabolites.

#### *Effects of dynamic exclusion and mass resolution in online LC-MS<sup>n</sup>.*

During online LC-MS<sup>n</sup> several parameters for optimizing the amount of metabolites fragmented or the fragmentation depth can be adapted in the Xcalibur acquisition software. The dynamic exclusion mode was used to automatically create a 20 seconds exclusion list of *m/z* values to be fragmented, in order to enable trapping of co-eluting lower abundant masses for MS<sup>n</sup> fragmentation. The use of this dynamic exclusion time markedly enlarged the coverage of tomato and Arabidopsis metabolites that were fragmented online, especially in the case of major metabolites co-eluting with lower abundant metabolites (Supplemental Table S2 in the Supporting Information). The short exclusion window of 20 s allows for repeated trapping and fragmenting of the same mass during LC-MS analysis, which is especially useful in the case of complex extracts containing closely eluting isomers.

Due to its significantly higher scanning rate, the nominal mass resolution Ion trap MS generated more fragmentation events within the same run than the Orbitrap FTMS. This can be illustrated by the number of MS<sup>2</sup> scan events of unique *m/z* values of 480, leading to an estimated amount of total MS<sup>2</sup> events of 650, while for MS<sup>3</sup> and MS<sup>4</sup> these numbers are 900 (1050) and 460 (550), respectively. The number of MS<sup>2</sup> and MS<sup>3</sup> spectra obtained with the Ion trap is significantly higher than obtained with the Orbitrap, while a significant amount of MS<sup>4</sup> spectra could be obtained as well.

To compare the Ion trap and the Orbitrap read-outs for reproducibility of the MS<sup>n</sup> spectral trees, we generated offline the fragmentation patterns of five different hexosides of naringenin and chalconaringenin (Supplemental Fig. 1A-D in the



Supporting Information). The absolute intensities of the parent ions and their fragments were generally about ten times higher for the Orbitrap than for the Ion trap (at a filling time of 100 ms), thus showing a higher sensitivity of the Orbitrap. In general, this led to a better signal-to-noise ratio for the Orbitrap data. Both the MS<sup>2</sup> and the MS<sup>3</sup> spectra were not identical between the two analyzers, as 2 minor fragments were only detected in the Orbitrap spectra. Nevertheless, the variations in relative intensities of the fragments detected by both analyzers were well within the criterion of 20%. Therefore, we conclude that spectral trees generated by the Orbitrap and the Ion trap match well.

*Offline MS<sup>n</sup> fragmentation in negative and positive ionization modes.*

The fractionation of the chromatographic peaks by the NanoMate robot collected sufficient amount of metabolite to perform in-depth offline MS<sup>n</sup> fragmentation, in both ionization modes. In fact, the 22  $\mu$ l fractions enabled over 90 minutes of nano-spraying. As an example, the offline spectral tree generation in both ionization modes of an *Arabidopsis* quercetin-diglycoside peak is shown in Fig. 4. The MS1 spectrum in negative mode was rather clean, while in positive mode another mass peak was detected within 1 Da of the target mass, i.e., within the mass selection window of the Ion trap. In negative mode, the MS<sup>2</sup> and MS<sup>3</sup> spectra showed the typical losses of one hexose and one deoxyhexose. The ionization was less efficient in positive mode than in negative mode and fragments of the co-fragmented mass peak were visible only in the positive MS<sup>2</sup> spectra. Nevertheless, the positive mode MS<sup>3</sup> spectra confirmed the presence of quercetin and the two sugar moieties. Upon comparison of the fragmentation spectra to reference data of differentially substituted quercetin glycosides,<sup>24</sup> most structural information was retrieved with the use of negative mode spectra. Based on the differences in the two negative mode MS<sup>3</sup> spectra displayed in Figure 4, showing the pronounced radical ion at  $m/z$  300 for the hexose loss, suggesting 3-O conjugation, and the  $m/z$  301 ion for the deoxyhexose loss, suggesting 7-O conjugation, we propose that this metabolite is a quercetin-3-O-hexose-7-O-deoxyhexose, most likely quercetin-3-O-glucoside-7-O-rhamnoside (see Figure



4 for structure). Also in tomato, the fragmentation in both ionization modes was key to discriminate between some isomers (e.g., Supplemental Figures 5 and 6 in the Supporting Information), thus illustrating the added value of performing offline MS<sup>n</sup> fragmentation.

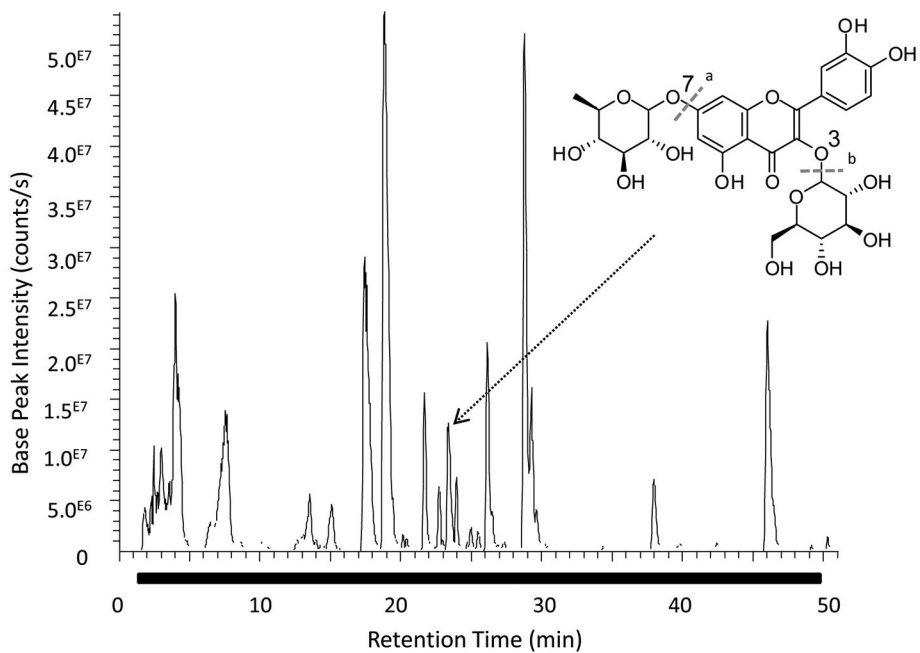
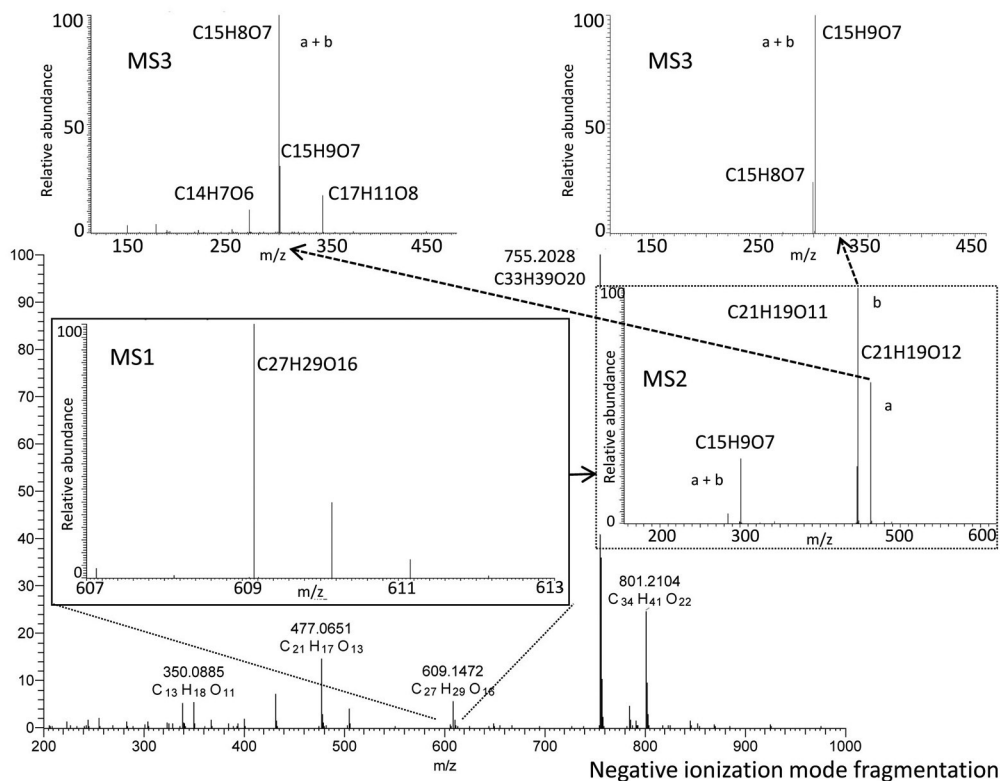
*MS<sup>n</sup> spectra enable discrimination of isomers.*

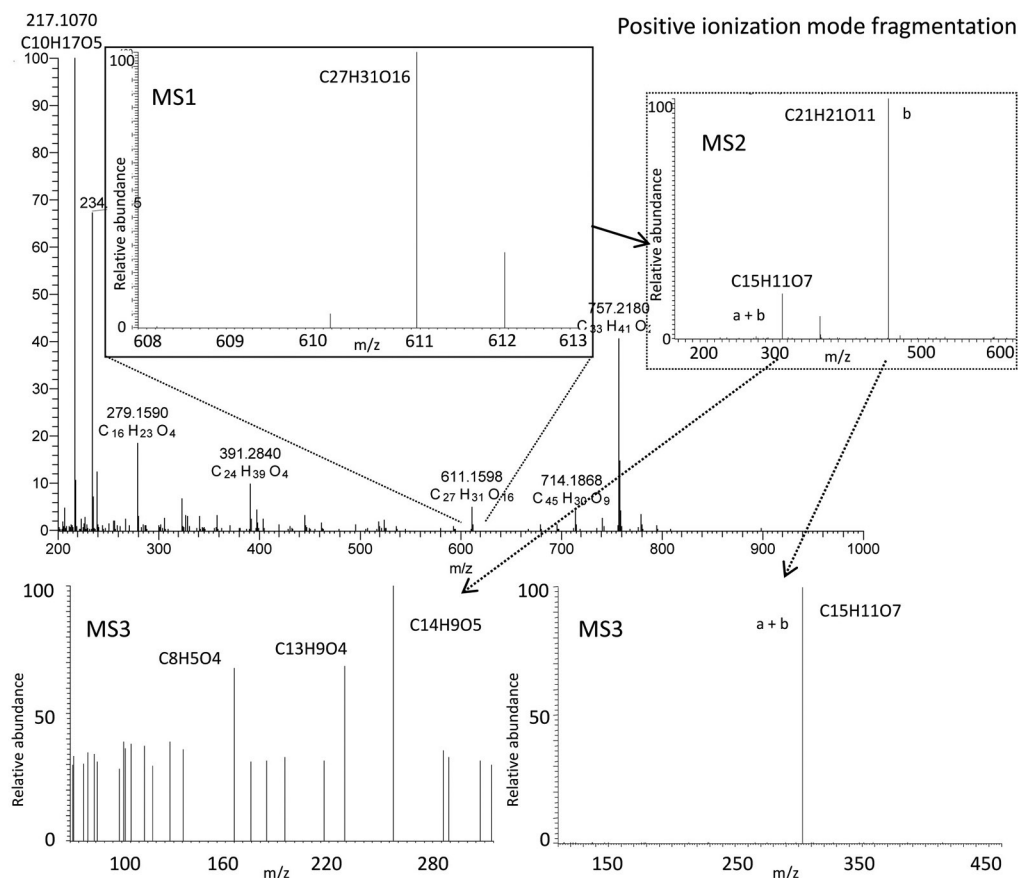
During full scan LC-Orbitrap FTMS profiling of tomato fruit, we detected several series of compounds with similar elemental formulas, i.e., isomers, typically within 1.5 ppm mass difference. We subsequently tested the ability of offline MS<sup>n</sup> fragmentation to differentiate between these isomers. An example of how the spectral data was analyzed and used for metabolite annotation is provided with Supplemental Figure 7 in the Supporting Information. Seven C<sub>21</sub>H<sub>22</sub>O<sub>10</sub> isomers were detected in the crude tomato extract and their MS<sup>2</sup> spectra (Supplemental Figure 1E in the Supporting Information) all share the common fragment C<sub>15</sub>H<sub>11</sub>O<sub>5</sub>, indicating a loss of C<sub>6</sub>H<sub>10</sub>O<sub>5</sub> (hexose). Subsequent fragmentation of the C<sub>15</sub>H<sub>11</sub>O<sub>5</sub> fragments resulted in identical MS<sup>3</sup> spectra, in both negative (Supplemental Figure 2 in the Supporting Information) and positive ionization mode (Supplemental Figure 3 in the Supporting Information), and based on the similarity with C<sub>15</sub>H<sub>12</sub>O<sub>5</sub> standards,<sup>24</sup> this led us to the annotation of these isomers as seven different naringenin/chalconaringenin (NG/CNG) – glycosides.

Following this robust MS<sup>n</sup> approach, we were able to discriminate all 7 selected isomers of phenylpropanoids (Supplemental Figure 4 in the Supporting Information) and flavonoids (Supplemental Figures 5-7 in the Supporting Information) in the crude tomato extract.

*LC-MS<sup>n</sup> enables large-scale annotation of metabolites.*

In order to test the power of MS<sup>n</sup> spectral trees in the characterization and annotation of metabolites present in crude and complex sample extracts, as detected by LC-MS based metabolomics approaches, we performed detailed analysis of MS<sup>n</sup> data generated from tomato fruit, focusing on phenolic





**Figure 4:** Part of LC-Orbitrap FT-MS chromatogram of *Arabidopsis* leaves, obtained in negative ionization mode. The chromatographic part used for fraction collection by the NanoMate robot for subsequent offline fragmentation is indicated by the black bar. Offline generated spectra in MS<sup>1</sup>, MS<sup>2</sup>, and MS<sup>3</sup> in both negative mode (upper spectra) and positive mode (lower spectra) of the same quercetin-diglycoside peak, eluting at 23.7 min, are presented. The putative structure based on the MS<sup>n</sup> spectra is displayed in the chromatogram. In the structure, the 3-O and 7-O positions are marked. The fragmentation events causing the major fragments are indicated by dashed grey lines in the structure marked with corresponding letters in both negative and positive mode fragmentation spectra.

compounds, and used an extract of *Arabidopsis* leaves to validate the MS<sup>n</sup> approach as generic tool. Supplemental Table 1 in the Supporting Information lists MS<sup>n</sup> data, generated either online or offline, from 99 selected metabolites in tomato and 28 metabolites in *Arabidopsis*. Of 16 metabolites detected in the tomato extract, the spectral data matched to fully annotated metabolites

reported in tomato literature,<sup>7</sup> while for 47 compounds only limited information has yet been provided, usually an elemental formula and retention time,<sup>7, 8, 10</sup> thus leaving many possible structures (MSI metabolite identification levels 3 and 4). The power of our LC-MS<sup>n</sup> approach is illustrated by the clear reduction of MSI level 4 identifications and the co-occurring shift towards level 2 and 3 identifications, as is shown for these 47 compounds in Supplemental Figure 8 in the Supporting Information. The other 36 compounds have not been reported before in tomato fruit and 21 thereof are even new in nature. A metabolite was considered novel if it was not present in any online metabolite repository (e.g. Scifinder, Dictionary of Natural Products, Kazusa, see also Suppl. Table 3). Based on their MS<sup>n</sup> spectra, these novel metabolites could already partially be annotated as conjugates, mainly glycosides, of caffeoylquinic acids and flavonoids, leading to MSI metabolite identification levels of 2 or 3. In addition, while the elemental formulas of 4 out of the 5 detected C<sub>30</sub>H<sub>28</sub>O<sub>13</sub> isomers have been described before in tomato,<sup>7</sup> MS<sup>n</sup> enabled us to discriminate between all of them and to annotate these isomeric series as being NG/CNG (C<sub>15</sub>H<sub>12</sub>O<sub>5</sub>) derivatives, revealing an C<sub>15</sub>H<sub>16</sub>O<sub>8</sub> substitution (m/z 324.0845). Moreover, seven novel hexose-substituted isomers of these C<sub>30</sub>H<sub>28</sub>O<sub>13</sub> NG/CNG derivatives with the elemental formula C<sub>36</sub>H<sub>38</sub>O<sub>16</sub> could be identified. Thus, our LC-MS<sup>n</sup> approach appeared a valuable tool to get better insight into the molecular structures of yet completely unknown or only partially annotated metabolites.

In order to verify the (putative) metabolite annotations based on the structural information as concluded from the MS<sup>n</sup> spectral data, four different polyphenols detected in *Arabidopsis* were purified and their structures were unambiguously established by NMR experiments. The compound predicted as being quercetin-3-*O*-glucoside-7-*O*-rhamnoside, based on its MS<sup>n</sup> spectral tree data (see above and Figure 4) was thus confirmed by NMR. The postulated structure of kaempferol-3-*O*-glucoside-7-*O*-rhamnoside, showing similar fragmentation behavior as its quercetin analogue in both ionization modes, was also confirmed by NMR. The negative mode MS<sup>2</sup> fragmentation patterns of two higher complex kaempferol-glycosides corresponded to a deoxyhexose conjugation at the 7-*O*

position, in view of both the absence of any radical fragment ions and the preferred fragmentation in negative ionization mode. The fragmentation of the resulting diglycoside moiety revealed clear differences in the number of observed fragments, corresponding to 1-2 and 1-6 linkages in their diglycoside moieties, respectively, based on their similarity to reference compounds.<sup>24</sup> The NMR measurements indeed identified these two metabolites as kaempferol-3-*O*-(2-*O*-rhamnosylglucoside)-7-*O*-rhamnoside and kaempferol-3-*O*-(6-*O*-glucosylglucoside)-7-*O*-rhamnoside. Thus, these 4 examples show that the structural prediction of metabolites based on MS<sup>n</sup> patterns was confirmed by unambiguous structural elucidation using NMR.

### *Discussion*

LC-MS based metabolomics approaches are widely used for profiling of complex biological extracts such as those from plants.<sup>1,2</sup> However, the identification of detected metabolites or marker compounds is still a bottleneck and although accurate mass LC-MS platforms are highly useful in rapid elemental formula determination, facilitating the metabolite annotation and identification process, the need for additional structural information from MS<sup>n</sup> experiments is evident from the numerous hits in different online databases for specific elemental formulas (Supplemental Table 3 and ref<sup>17</sup>). Here, we applied a robust, accurate mass based MS<sup>n</sup> spectral tree approach for the characterization and annotation of metabolites detected by LC-MS analysis of crude plant extracts, with the use of a LTQ/Orbitrap hybrid mass spectrometer, enabling both partial tree generation by online MS<sup>n</sup> and in-depth tree generation offline using the NanoMate fractionation/injection-robot. While the time available for online MS<sup>n</sup> of compounds is restrained by their chromatographic peak widths, offline MS<sup>n</sup> after LC-MS fractionation does not depend on chromatographic conditions. In addition, the offline MS<sup>n</sup> approach with the chip-based nano-spray source consumes only a few nanoliters per minute and thus in-depth fragmentation experiments can be performed in both ionization modes and specific reagents can be added to enhance ionization or fragmentation, if needed. The MS<sup>n</sup> approach enabled us

to recognize common substructures shared between metabolites and to describe their fragmentation patterns in more detail as compared to other (accurate mass) MS/MS platforms.<sup>7, 8, 17</sup> Within one sample run, the online LC-MS<sup>n</sup> approach produced high-density structural information data of many metabolites present in the crude plant extracts.

Both online and offline MS<sup>n</sup> fragmentation of metabolites resulted in robust MS<sup>n</sup> patterns and reproducible spectral trees derived thereof (Figure 3). Moreover, the partial spectral trees of metabolites obtained by online LC-MS<sup>n</sup> matched well with both their offline in-depth spectral trees and with those obtained from reference compounds (Figure 3C). These results indicate that the MS<sup>n</sup> fragmentation is independent of the solvent and LC conditions applied. In addition, the spectral trees were also stable both upon changes in either compound concentration, as observed from sequential spectra of a chromatographic peak (Figure 3A) and from dilution series of standards, and MS<sup>n</sup> spectra obtained at different normalized collision energies of the Ion trap,<sup>24</sup> and over prolonged time between spectral tree generations (Figure 3B and ref<sup>24</sup>). Therefore, the structural information that can be extracted from MS<sup>n</sup> patterns, i.e., both the presence of unique mass fragments and the relative intensities of fragment ions, including radical fragments,<sup>24</sup> can be used for the determination of compound structure and substructures, by matching the spectral trees of unknown metabolites with that of reference compounds. Moreover, structural relations between unknown compounds, including series of isomers, can be deduced from their spectral similarities at a specific MS<sup>n</sup> level (e.g. Fig. 1 and Suppl. Table S1 in the Supporting Information).

#### *Factors influencing online LC-MS<sup>n</sup> fragmentation.*

During online LC-MS<sup>n</sup> several factors and parameters influence the coverage of fragmented metabolites and the tree size. Firstly, during online LC-MS<sup>n</sup>, the fragmentation depth of metabolites is restricted by the chromatographic peak width, which is mostly influenced by the chromatographic conditions applied and the metabolite abundance. The depth of online spectral trees also depends on the fragmentation efficiency and the ionization efficiency of the metabolite and

its fragments. Secondly, the use of dynamic exclusion is key in unbiased online MS<sup>n</sup> approaches (Supplemental Table S2 in the Supplemental Information), as high abundant parent ions and their accompanying mass features, like isotopes, in-source fragments, doubly charged species and other adducts, can otherwise prevent the fragmentation of a lower abundant co-eluting compound. Lastly, as the scan times are slightly variable for the Orbitrap platform, due to the changing metabolite abundances in combination with the use of the automatic gain controller (AGC), online MS<sup>n</sup> fragmentation is not always guaranteed for low abundant metabolites. The faster scanning rate of the Ion trap, as compared to the Orbitrap, enables fragmentation of more metabolites as well as deeper MS<sup>n</sup> spectra (Supplemental Table S2 in the Supplemental Information). The spectra obtained by the Ion trap and the Orbitrap, including relative intensities of the fragments, were identical, except for two minor fragments observed in the Orbitrap (Supplemental Figure 1A-D in the Supporting Information). Unmistakably, high mass resolution is important for rapid determination of the elemental formula of detected ions and fragments. For example, the discrimination of C<sub>13</sub>H<sub>13</sub>O<sub>9</sub> (m/z 313.0560) and C<sub>17</sub>H<sub>13</sub>O<sub>6</sub> (m/z 313.0712) fragment ions observed for the NG/CNG-hexosides and dihexosides underlines the benefit of working with accurate mass (Supplemental Figures 1E and 7 in the Supporting Information). However, once high mass resolution Orbitrap-MS<sup>n</sup> spectra of the same compound are already available for comparison, the nominal mass Ion trap-MS<sup>n</sup> spectra can subsequently be used to rapidly annotate these metabolites in crude extracts, making use of the fast scanning rate of the Ion trap.

*Metabolite discrimination and annotation using MS<sup>n</sup> fragmentation.*

Polyphenols<sup>27, 28</sup> and glucosinolates<sup>29, 30</sup> have frequently been fragmented using Ion trap-mediated MS<sup>n</sup> fragmentation. However, the robustness of the resulting fragmentation patterns, and thus the reproducibility of the MS<sup>n</sup> spectral trees, has not been taken into account in these previous studies,<sup>24</sup> thereby hampering exact matching of fragmentation spectra. Moreover, most MS<sup>n</sup> fragmentation experiments have previously been carried out at nominal mass resolution.

Nevertheless, MS<sup>n</sup> experiments have surely increased our knowledge about typical ion fragmentation paths in both polyphenols<sup>27, 31, 32</sup> and glucosinolates,<sup>29</sup> thereby defining metabolite fragmentation rules and facilitating metabolite annotation.

By definition, isomeric compounds have the same elemental formula, and thus accurate mass, and crude (plant) extracts may contain different isomer series. However, with the use of MS<sup>n</sup> we showed that isomers from different plant compound classes, including phenylpropanoids and flavonoids, could be discriminated. This discriminative power was frequently different between negative and positive ionization modes, underlining the importance of generating spectral trees in both ionization modes (Supplemental Figures 1-7 in the Supporting Information and ref<sup>33</sup>). As all *Arabidopsis* and tomato flavonoid isomers studied, except the tomato chalconaringenin/naringenin couple, could be differentiated based on their MS<sup>n</sup> spectral trees generated in either negative or positive ionization mode, thereby creating a unique fingerprint, we conclude that the proposed accurate mass LC-MS<sup>n</sup> approach is of great help in discriminating and annotating isomeric metabolites within and between crude extracts.

The application of the MS<sup>n</sup> spectral tree approach in LC-MS analysis of complex plant extracts allowed the characterization and annotation of several series of biosynthetically related metabolites in tomato and *Arabidopsis*, of which some were new or yet only partially annotated (Supplemental Table S1 in the Supporting Information). For tomato fruit, we describe 36 new compounds of which 21 were unknown so far. For example, a C<sub>3</sub>H<sub>6</sub>O<sub>3</sub>S substitution of selected flavonoids was observed three times (Supplemental Table S1 in the Supporting Information). So far, sulphur containing substitutions on flavonoids other than C<sub>3</sub>H<sub>7</sub>O<sub>2</sub>NS (most likely cysteine) have not been reported in tomato. A possible candidate for the C<sub>3</sub>H<sub>6</sub>O<sub>3</sub>S substitution is methylthio-acetic acid. These results illustrate that LC-MS<sup>n</sup> approaches can provide structural information about both novel and earlier observed, but not yet (fully) annotated (mainly MSI identification levels 3 and 4), metabolites in crude extracts, enabling specification of the basic structures in conjugated metabolites, as well as of substitutions and substructures that



metabolites share.

The robustness of the off-line MS<sup>n</sup> fragmentation patterns and their similarity with patterns from the online fragmentation of selected metabolites suggest that ESI based fragmentation rules from literature can be directly applied to metabolites present in crude extracts.<sup>31, 32</sup> For instance, our fragmentation of the intact glucosinolates in Arabidopsis gave similar results compared to earlier Ion trap fragmentation studies,<sup>29, 30</sup> indicating the reproducibility of MS<sup>n</sup> spectra even on different Ion trap based mass spectrometry platforms. The examples of substitutions on flavonoids (Figures 1 and 4) demonstrate that, in negative ionization mode, substitutions on the 7-O position fragment more easily and generated different spectral trees as compared to 3-O substitutions.<sup>24, 34</sup> The predicted structures of four annotated metabolites, based on MS<sup>n</sup>, were subsequently confirmed by NMR, indicating that MS<sup>n</sup> data can provide fragmentation rules that can be used in metabolite identification in practice. In addition, previously identified metabolites can be easily detected as substructures within more complex compounds (Figure 1). Although complex rearrangements can occur during fragmentation in aromatic molecules like polyphenols, we observed highly reproducible MS<sup>n</sup> fragmentation spectra for this compound class in both ionization modes, indicating the ability of using MS<sup>n</sup> fragmentation patterns and spectral trees for library matching. Furthermore, an increasing amount of spectral trees generated from completely structurally elucidated metabolites (i.e., MSI identification level 1) will lead to a better insight into ESI-mediated fragmentation and may provide generic fragmentation rules that can help in MS<sup>n</sup> spectra interpretation and metabolite identification.

In order to decrease the time for identifying metabolites by assigning structural elements, dedicated software is needed that efficiently extracts the relevant and discriminative information from the fragmentation data and that enables scientists to automatically match fragmentation and structural aspects of molecules, e.g., comparable to the NIST library (NIST/EPA/NIH Mass Spectral Library, NIST 08) for electron impact GC-MS. This can eventually lead to the untargeted systematic MS<sup>n</sup> analysis of 5 sec fractions of crude extracts within several days.

At the moment, there is no software available that can automatically process, visualize and compare accurate mass based  $MS^n$  spectra in an unbiased manner. Although the spectral tree concept as suggested by Sheldon et al.<sup>25</sup> is quite similar to ours, we used accurate mass instead of nominal mass  $MS^n$  data, focused on its reproducibility aspects and applied it on a larger set of reference compounds<sup>24</sup> and tested and optimized its applicability for annotating yet known and unknown compounds in crude extracts (present work). In addition, while the MassFrontier software can visualize spectral trees,<sup>25</sup> it currently has substantial drawbacks, like the inclusion of noise peaks due to the fact that masses instead of elemental formula are used.<sup>24</sup> Therefore, new software tools are currently being developed within the Netherlands Metabolomics Centre, e.g. for peak picking and elemental formula assignments making use of the high resolution  $MS^n$  spectral tree data, as well as spectral tree viewing and library searching, which are key steps in automation of metabolite identification based on  $MS^n$  data. While the current methodology was tested on an Ion trap - Orbitrap mass analyzer, the NanoMate robot may as well be coupled to other types of mass analyzers such as (Q-)TOF and FT-ICR machines, in order to generate fragmentation spectra from large series of metabolites at different platforms. This approach can provide relevant information about fragmentation patterns and product ions that are either in common or unique to the type of mass analyzer used. Recently, it was shown that spectral trees computed from MS/MS data obtained at different collision energies can be in good agreement with  $MS^n$  spectra,<sup>18</sup> suggesting that differently generated mass spectral databases can be transferred *in silico* and combined for automated searching of unknown metabolites. All these developments indicate that suitable software for automated processing of accurate mass  $MS^n$  data will very likely become available in the nearby future. Once this software is available,  $MS^n$  spectral trees can provide a rapid large-scale identification tool, thereby decreasing the time needed for metabolite profiling. In our study, we present an approach that generates, both in an online and offline manner, the robust  $MS^n$  patterns from metabolites in complex crude extracts needed as input for this dedicated  $MS^n$  software tools and databases.

## *Conclusions*

Both online LC-MS<sup>n</sup> and offline MS<sup>n</sup> generated spectral trees and their fragmentation patterns, generated by the Ion trap - Orbitrap system, may be used as reproducible metabolite fingerprints. As shown using complex extracts from plants, the spectral tree approach can discriminate closely related molecules like isomeric compounds, and specify substructures that are in common between different metabolites and their conjugates. In tomato, detailed analysis of MS<sup>n</sup> fragmentation patterns of phenolic compounds, led to the characterization and annotation of 21 phenolic compounds not reported before in literature or available metabolite databases. In addition, with an increasing amount of annotated metabolites that are completely structurally elucidated, more ESI-Ion trap based MS<sup>n</sup> fragmentation rules can be derived, thus further facilitating metabolite identification by MS in the future. We therefore believe that the MS<sup>n</sup> spectral tree method is a very powerful tool in the annotation of metabolites in crude extracts, as well as in structural elucidation of unknown and partly unknown metabolites by comparing spectral trees to each other and to reference data. Due to the observed robustness, reproducibility and discriminative power, MS<sup>n</sup> databases and an integrated automated workflow for rapid metabolite identification using MS<sup>n</sup> spectra trees is foreseen in the nearby future.

## *Acknowledgements*

This research was granted by the Netherlands Metabolomics Centre (JJvdH, JV, RCHdV) and the Centre for BioSystems Genomics (RJB, RCHdV), both of which are part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. The authors specifically thank Rik Kooke and Joost Keurentjes (Laboratory of Plant Physiology, WUR) for growing, harvesting and grinding the Arabidopsis ecotype population, Harry Jonker and Bert Schipper (PRI, WUR) for technical assistance, Piotr Kasper, Miguel Rojas, and Rob Vreeken (LACDR, Leiden University) for valuable discussions on MS<sup>n</sup> fragmentation and Velitchka Mihaleva (Laboratory of Biochemistry, WUR) for fruitful comments on the manuscript.

## Supporting Information Available

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>. The supplemental figures are also part of this thesis, pages 107-116, and the supplemental Excel file is available at <http://edepot.wur.nl/216854>

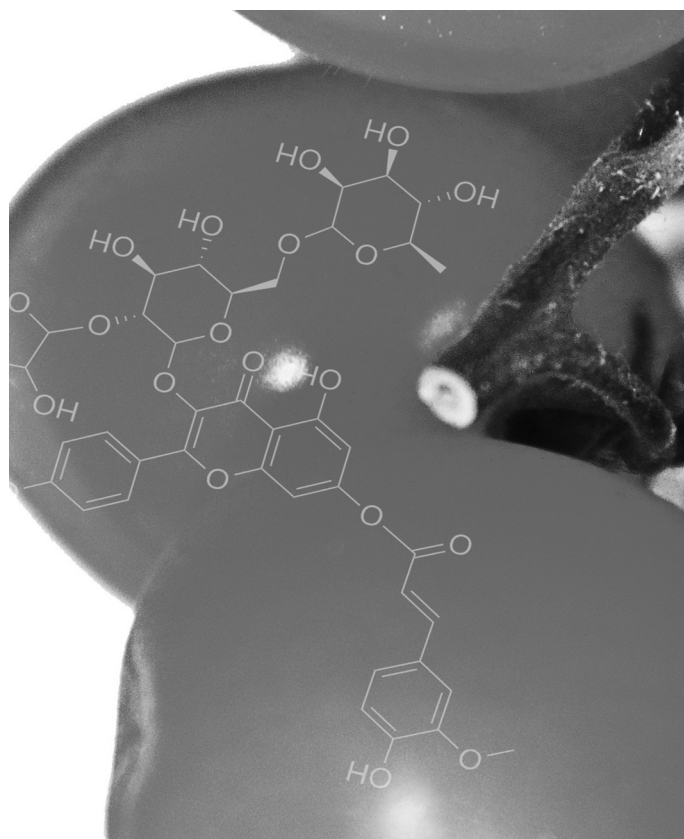
## References

1. E. Werner, J. F. Heilier, C. Ducruix, E. Ezan, C. Junot and J. C. Tabet, *J. Chromatogr., B*, 2008, **871**, 143-163.
2. W. B. Dunn, *Phys. Biol.*, 2008, **5**, art. no. 011001.
3. T. Kind and O. Fiehn, *Bmc Bioinf.*, 2007, **8**, art. no. 105.
4. S. Moco, R. J. Bino, R. C. H. De Vos and J. Vervoort, *Tr. Anal. Chem.*, 2007, **26**, 855-866.
5. M. Bedair and L. W. Sumner, *Tr. Anal. Chem.*, 2008, **27**, 238-250.
6. C. Böttcher, E. v. Roepenack-Lahaye, J. Schmidt, C. Schmotz, S. Neumann, D. Scheel and S. Clemens, *Plant Physiol.*, 2008, **147**, 2107-2110.
7. Y. Iijima, K. Suda, T. Suzuki, K. Aoki and D. Shibata, *J. Jap. Soc. Hortic. Sci.*, 2008, **77**, 94-102.
8. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and R. C. H. De Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.
9. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211-221.
10. M. Gómez-Romero, A. Segura-Carretero and A. Fernández-Gutiérrez, *Phytochemistry*, 2010, **71**, 1848-1864.
11. S. Moco, E. Capanoglu, Y. Tikunov, R. J. Bino, D. Boyacioglu, R. D. Hall, J. Vervoort and R. C. H. De Vos, *J. Exp. Bot.*, 2007, **58**, 4131-4146.
12. A. Vallverdú-Queralt, O. Jáuregui, A. Medina-Remón, C. Andrés-Lacueva and R. M. Lamuela-Raventós, *Rapid Commun. Mass Spectrom.*, 2010, **24**, 2986-2992.
13. E. v. Roepenack-Lahaye, T. Degenkolb, M. Zerjeski, M. Franz, U. Roth, L. Wessjohann, J. Schmidt, D. Scheel and S. Clemens, *Plant Physiol.*, 2004, **134**, 548-559.
14. T. Tohge, K. Yonekura-Sakakibara, R. Niida, A. Watanabe-Takahashi and K. Saito, *Pure Appl. Chem.*, 2007, **79**, 811-823.
15. F. Matsuda, K. Yonekura-Sakakibara, R. Niida, T. Kuromori, K. Shinozaki and K. Saito, *Plant J.*, 2009, **57**, 555-577.
16. R. Slimestad and M. Verheul, *J. Science Food Agric.*, 2009, **89**, 1255-1270.
17. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23-60.
18. F. Rasche, A. Svatos, R. K. Maddula, C. Böttcher and S. Böcker, *Anal. Chem.*, 2011, **83**, 1243-1251.
19. F. Ferreres, M. Taveira, D. M. Pereira, P. Valentão and P. B. Andrade, *J. Agr. Food Chem.*, 2010, **58**, 2854-2861.
20. M. Francisco, D. A. Moreno, M. E. Cartea, F. Ferreres, C. García-Viguera and P. Velasco, *J. Chromatogr., A*, 2009, **1216**, 6611-6619.

21. A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat and S. Horning, *Anal. Chem.*, 2006, **78**, 2113-2120.
22. A. Makarov and M. Scigelova, *J. Chromatogr., A*, 2010, **1217**, 3938-3945.
23. M. Scigelova and A. Makarov, *Proteomics*, 2006, **1**, 16-21.
24. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino, J. Beekwilder and C. H. R. de Vos, *Anal. Chem.*, 2011, **83**, 409-416.
25. M. T. Sheldon, R. Mistrik and T. R. Croleya, *J. Am. Mass Spectrom. Soc.*, 2009, **20**, 370-376.
26. C. H. R. de Vos, S. Moco, A. Lommen, J. J. Keurentjes, R. J. Bino and R. D. Hall, *Nat. Prot.*, 2007, **2**, 778-791.
27. V. Vukics and A. Guttman, *Mass Spectrom. Rev.*, 2010, **29**, 1-12.
28. J. L. Wolfender, P. Waridel, K. Ndjoko, K. R. Hobby, H. J. Major and K. Hostettmann, *Analisis*, 2000, **28**, 895-906A.
29. S. J. Rochfort, V. C. Trenerry, M. Imsic, J. Panozzo and R. Jones, *Phytochemistry*, 2008, **69**, 1671-1679.
30. S. Millán, M. C. Sampedro, P. Gallejones, A. Castellón, M. L. Ibargoitia, M. A. Goicolea and R. J. Barrio, *Anal. Bioanal. Chem.*, 2009, **394**, 1661-1669.
31. S. J. Rochfort, M. Imsic, R. Jones, V. C. Trenerry and B. Tomkins, *J. Agr. Food Chem.*, 2006, **54**, 4855-4860.
32. H. Olsen, K. Aaby and G. I. A. Borge, *J. Agr. Food Chem.*, 2009, **57**, 2816-2825.
33. M. N. Clifford, S. Knight and N. Kuhnert, *J. Agr. Food Chem.*, 2005, **53**, 3821-3832.
34. Y. L. Ma, F. Cuyckens, H. Van den Heuvel and M. Claeys, *Phytochem. Anal.*, 2001, **12**, 159-165.

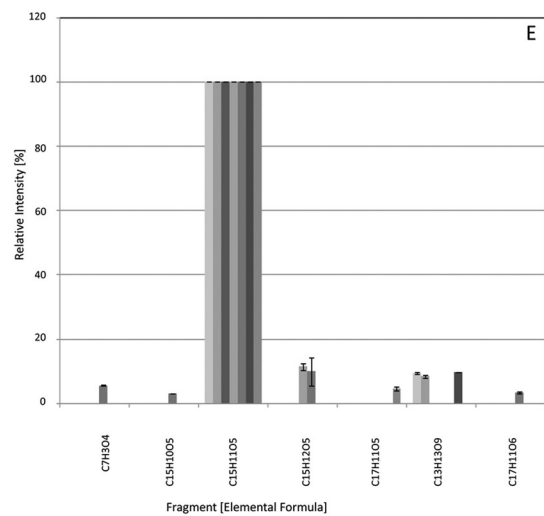
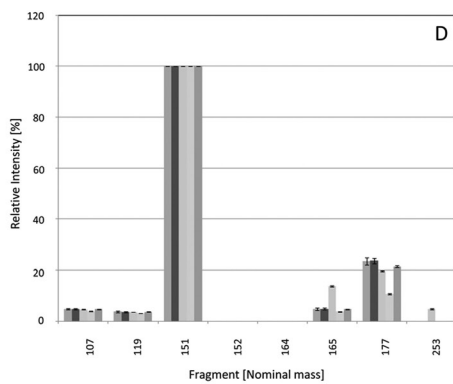
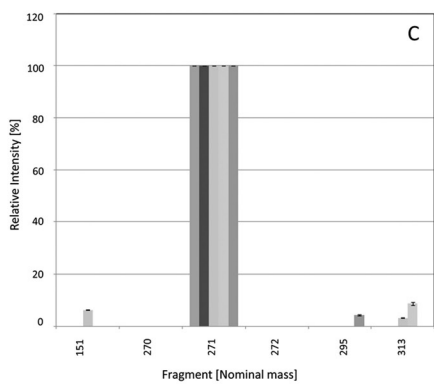
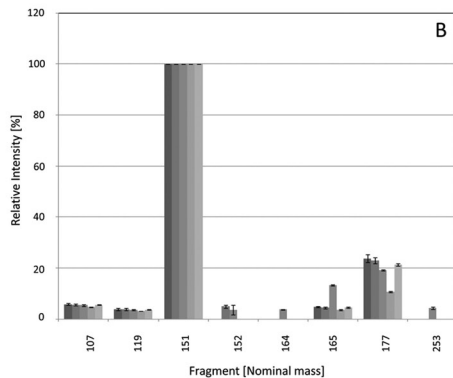
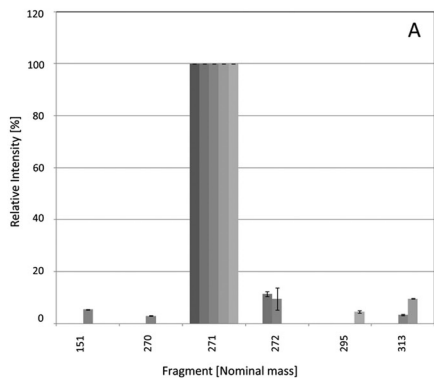


Supporting information to *Chapter 3*:  
'Spectral trees as a robust annotation tool in  
LC-MS based metabolomics'



Justin J.J. van der Hooft, Jacques Vervoort, Raoul J. Bino,  
and Ric C.H. de Vos

This Supporting Information was published online with the research article in *Metabolomics*, 2012, Volume 8 (4), pp 691–703, DOI: 10.1007/s11306-011-0363-7



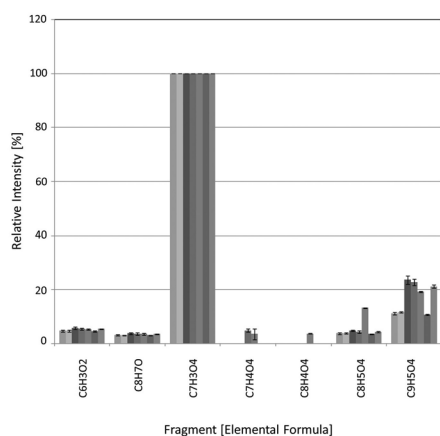


**Supplemental Figure 1 (previous page):** Offline MS<sup>n</sup> fragmentations, recorded in negative mode, of 5 tomato naringenin/chalconaringenin hexoses in the Orbitrap (A,B) and the Ion trap (C,D). E) Offline negative mode accurate mass MS<sup>2</sup> spectra of 7 naringenin/chalconaringenin-hexose metabolites in tomato.

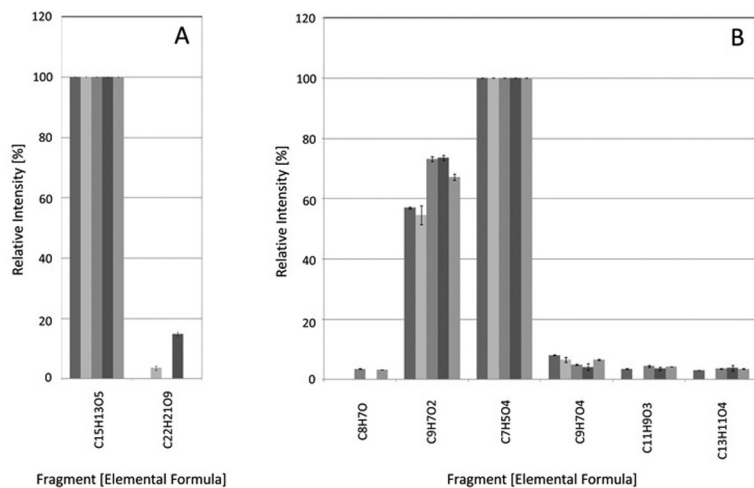
A) MS<sup>2</sup> and C) MS<sup>3</sup> (271 fragment) spectra in the Orbitrap of five naringenin/chalconaringenin-hexoside isomers that eluted at retention time 31.47 (dark grey), 31.91 (grey), 33.24 (mid grey), 35.74 (light grey) and 36.28 min (lightest grey). For the purpose of comparing the fragmentation read-outs of both mass spectrometers, the Orbitrap accurate mass data of spectra A and C were transferred into nominal masses.

B) MS<sup>2</sup> and D) MS<sup>3</sup> (271 fragment) spectra in the Ion Trap of (the same) five naringenin/chalconaringenin-hexosides, at retention time 31.47 (mid grey), 31.91 (black), 33.24 (lighter grey), 35.74 (lightest grey), and retention time 36.28 (grey). E) Offline MS<sup>2</sup> accurate mass spectra in negative ionization mode are shown of seven tomato fruit naringenin/naringenin chalcone-hexose metabolites eluting at retention time 25.48 (mid grey), 26.48 (light grey), 31.47 (dark grey), 31.91 (light grey), 33.24 (mid grey), 35.74 (black), and 36.28 (mid grey).

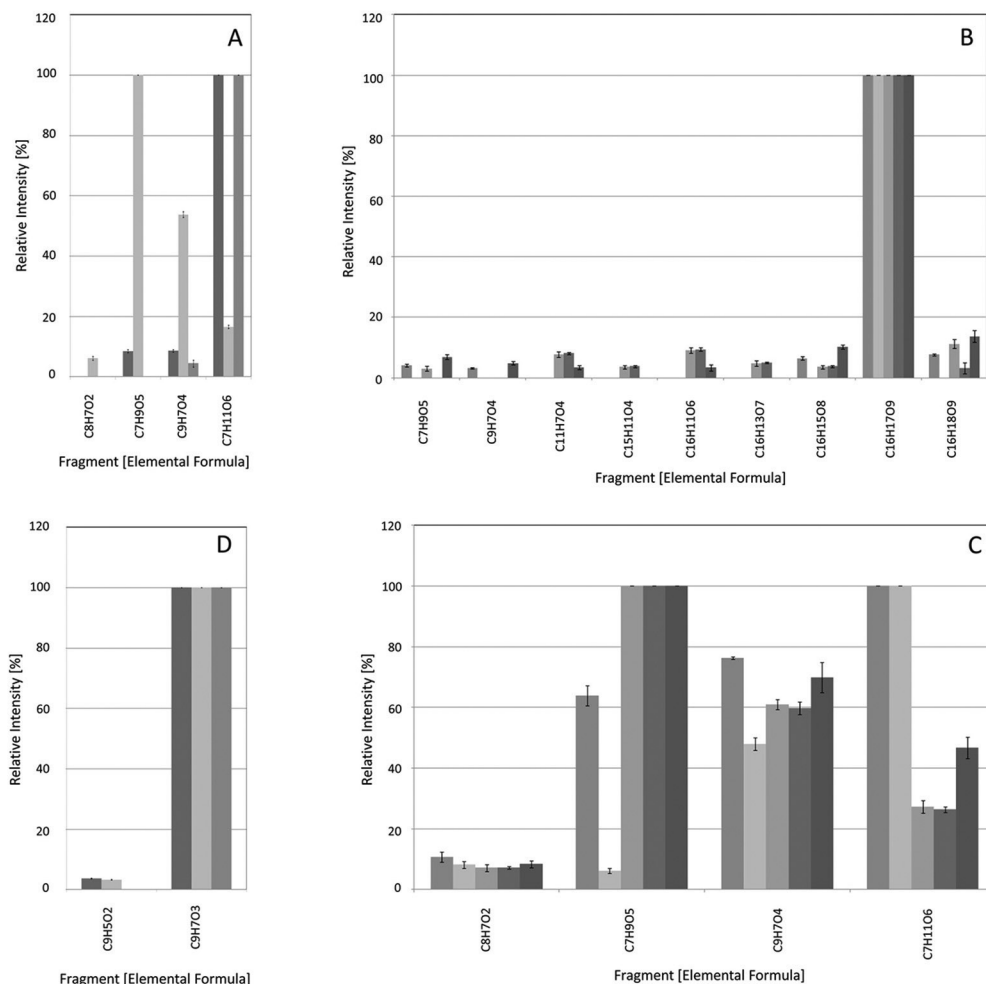
MS<sup>n</sup> data represent averaged spectra ( $n = 6$ ) with error bars representing two times the standard deviation.



**Supplemental Figure 2:** MS<sup>3</sup> fragmentation of naringenin/chalconaringenin hexoses in negative ionization mode. Bars represent averaged ( $n=6$ ) offline MS<sup>3</sup> accurate mass spectra of seven naringenin/chalconaringenin-hexoses eluting at retention time 25.48 (light grey), 26.48 (lightest grey), 31.47 (dark grey), 31.91 (mid grey), 33.24 (light grey), 35.74 (grey) and 36.28 min (mid grey). Error bars represent two times the standard deviation.



**Supplemental Figure 3:** MS<sup>n</sup> fragmentation of naringenin/chalconaringenin hexoses in positive ionization mode. Bars represent averaged ( $n=6$ ) A) offline MS<sup>2</sup> and B) MS<sup>3</sup> ( $C_{15}H_{13}O_5$  fragment,  $m/z$  273) accurate mass spectra of five naringenin/chalconaringenin-hexoside based isomers, at retention time 31.47 (dark grey), 31.91 (light grey), 33.24 (mid grey), 35.74 (black), and retention time 36.28 (mid grey). Error bars represent two times the standard deviation.



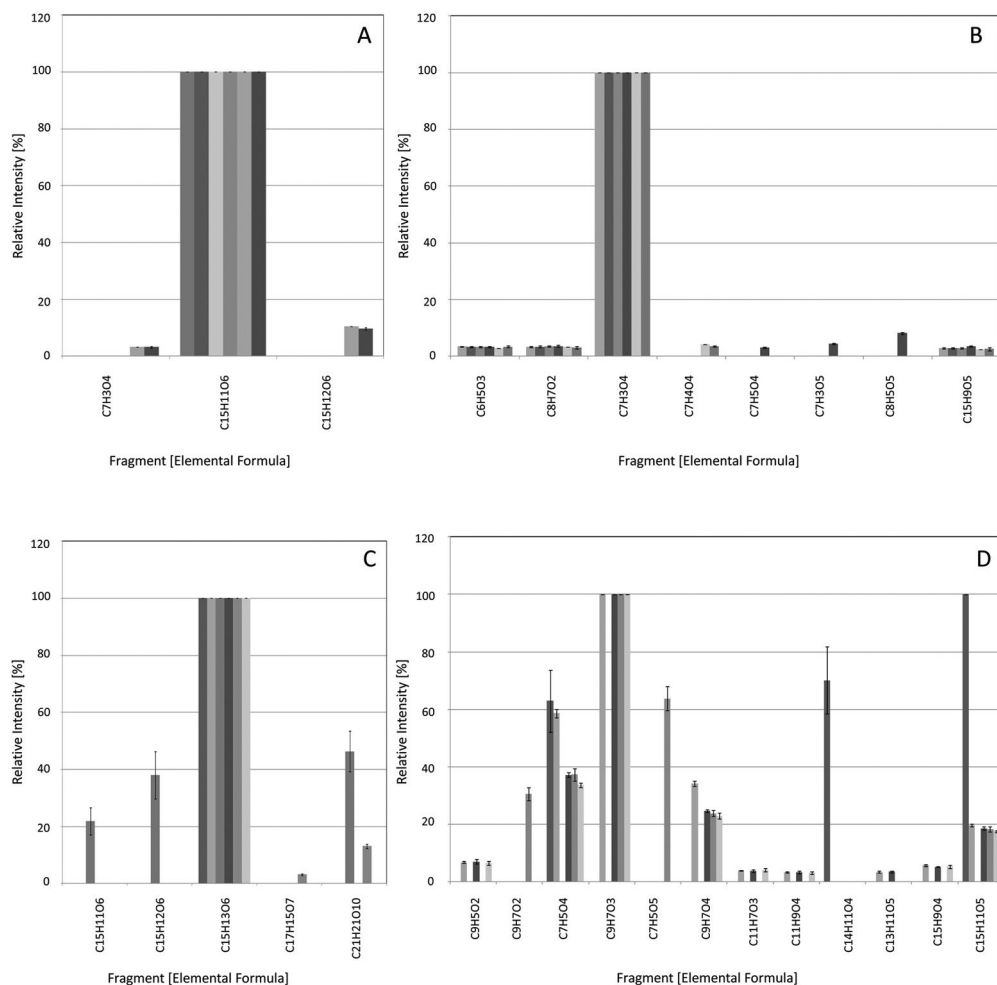
**Supplemental Figure 4:** Accurate mass MS<sup>n</sup> fragmentation of caffeoylquinic and dicaffeoylquinic acids in tomato fruit. A-C: negative ionization mode, D positive ionization mode.

A) offline MS<sup>2</sup> spectra of three caffeoylquinic acids eluting at retention times 16.19 (dark grey), 17.12 (light grey), and 19.18 (mid grey, n=4);

B) offline MS<sup>2</sup> and C) MS<sup>3</sup> (C<sub>16</sub>H<sub>17</sub>O<sub>9</sub> fragment, m/z 353) spectra of four C<sub>25</sub>H<sub>23</sub>O<sub>12</sub> ions at retention time 30.13 (dark grey), 30.85 (light grey), two consecutive scans at 33.04 (mid grey and darkest grey), and 42.06 min (black);

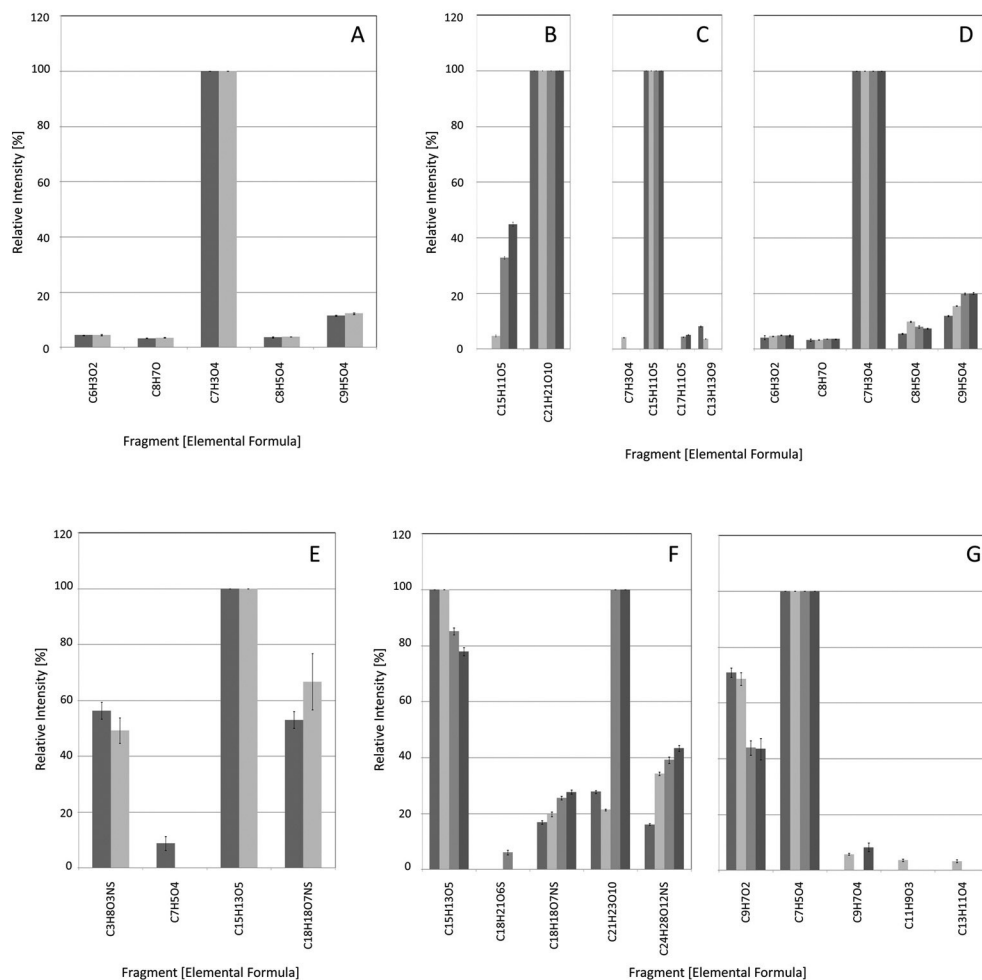
D) positive mode MS<sup>2</sup> spectra of caffeoylquinic acids eluting at retention time 16.19 (dark grey), 17.12 (light grey), and 19.18 (mid grey).

MS<sup>n</sup> data represent averaged spectra (n = 6, unless stated differently) with error bars representing two times the standard deviation.



**Supplemental Figure 5:** Offline accurate mass MS<sup>n</sup> fragmentation of C<sub>15</sub>H<sub>12</sub>O<sub>6</sub> based flavonoid glycosides.

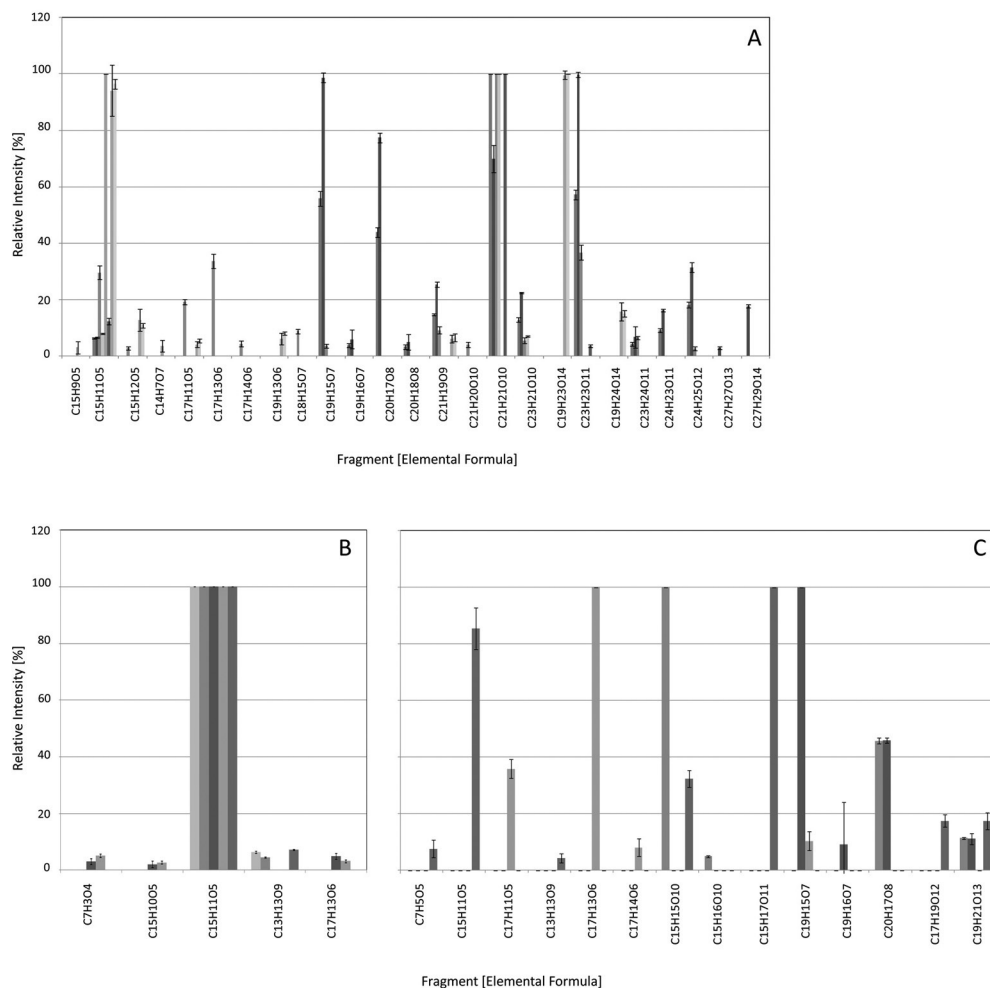
*A and B:* negative mode offline MS<sup>2</sup> (A) and MS<sup>3</sup> (C<sub>15</sub>H<sub>11</sub>O<sub>6</sub> fragment) (B) accurate mass spectra of five C<sub>15</sub>H<sub>12</sub>O<sub>6</sub> based glycosides at retention time 27.06 two consecutive scans (light grey and dark grey), 30.99 (mid grey), 32.76 (darkest grey), and two consecutive scans at retention time 34.98 (grey and black). *C and D:* positive mode offline MS<sup>2</sup> (C) and MS<sup>3</sup> (C<sub>15</sub>H<sub>13</sub>O<sub>6</sub> fragment) (D) spectra of six C<sub>15</sub>H<sub>12</sub>O<sub>6</sub> based glycosides at retention time 18.91 (dark grey), 27.06 (light grey), 30.99 (mid grey), 32.25 (darkest grey), 32.76 (grey), and at retention time 34.98 (lightest grey). MS<sup>n</sup> data represent averaged spectra (n = 6) with error bars representing two times the standard deviation.



**Supplemental Figure 6:** Offline accurate mass MS<sup>n</sup> fragmentation of naringenin/chalconaringenin aglycones and naringenin/chalconaringenin-hexoses with C<sub>3</sub>H<sub>7</sub>O<sub>2</sub>NS adducts.

The negative mode (A) and positive mode (B) offline MS<sup>2</sup> spectra of two naringenin/chalconaringenin- C<sub>3</sub>H<sub>7</sub>O<sub>2</sub>NS metabolites at retention time 23.75 (dark grey) and 24.32 (light grey); and the negative mode MS<sup>2</sup> spectra (C), MS<sup>3</sup> spectra (C<sub>21</sub>H<sub>21</sub>O<sub>10</sub> fragment) (D), and MS<sup>4</sup> spectra (C<sub>21</sub>H<sub>21</sub>O<sub>10</sub> - C<sub>15</sub>H<sub>11</sub>O<sub>6</sub> fragment) (E), as well as the positive mode MS<sup>2</sup> spectra (F), and MS<sup>3</sup> spectra (C<sub>15</sub>H<sub>13</sub>O<sub>6</sub> fragment) (G) of four naringenin/chalconaringenin-hexose- C<sub>3</sub>H<sub>7</sub>O<sub>2</sub>NS metabolites at retention time 17.47 (dark grey), 18.37 (light grey), 18.85 (grey), and retention time 19.52 (darkest grey).

MS<sup>n</sup> data represent averaged spectra ( $n = 6$ ) with error bars representing two times the standard deviation.

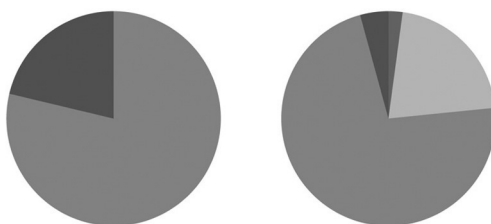


**Supplemental Figure 7:** Offline accurate mass  $MS^n$  fragmentation of naringenin/chalconaringenin-dihexose isomers. (A) The negative mode  $MS^2$  spectra of 7 naringenin/chalconaringenin-dihexoses, at retention time 17.60 (mid grey,  $n=4$ ), 18.37 (darkest grey), 22.29 (grey), 25.68 (lightest grey,  $n=4$ ), 26.85 (light grey), 27.57 (dark grey), and two consecutive scans of retention time 35.10 (grey and light grey,  $n=4$ ). (B) The negative mode  $MS^3$  spectra ( $C_{21}H_{21}O_{10}$  fragment) of 5 naringenin/chalconaringenin-dihexoses, at retention time 17.60 (dark grey,  $n=4$ ), 18.37 (light grey), 22.29 (grey), 25.68 (mid grey,  $n=4$ ), and retention time 27.57 (grey). (C) The negative mode  $MS^3$  spectra ( $C_{21}H_{23}O_{10}$  fragment) of 4 naringenin/chalconaringenin-dihexoses, at retention time 17.60 (mid grey,  $n=4$ ), 18.37 (darkest grey), 22.29 (light grey), and retention time 27.57 (grey).

$MS^n$  data represent averaged spectra ( $n = 6$ , unless stated differently) with error bars representing two times the standard deviation.

*Example of MS<sup>n</sup> data analysis and metabolite identification belonging to Supplemental Figure 7:*

The MS<sup>n</sup> fragmentation behavior of naringenin/chalconaringenin (NG/CNG)-dihexosides was similar to their monohexoside analogues (Supplemental Figure 7 in the Supporting Information). Some NG/CNG-dihexosides generated unique fragmentation spectra in MS<sup>2</sup>, e.g. the isomer eluting at RT 18.37 min showed typical losses of H<sub>2</sub>O, suggesting a 1,2-linkage between the two hexose moieties.<sup>22, 23</sup> Comparison of the MS<sup>2</sup> data with publicly available accurate mass databases, e.g. the KOMICS database ([www.webs2.kazusa.or.jp/komics/index.php](http://www.webs2.kazusa.or.jp/komics/index.php))<sup>7</sup> confirmed this C<sub>27</sub>H<sub>32</sub>O<sub>15</sub> isomer to be naringin, i.e. naringenin-7-O-glucoside-1,2-glucoside.<sup>7, 8</sup> The MS<sup>3</sup> spectra from the C<sub>21</sub>H<sub>21</sub>O<sub>10</sub> (m/z 433, loss of one hexose moiety) fragments observed for all 5 NG/CNG-dihexosides generated similar fragmentation patterns, with the C<sub>15</sub>H<sub>11</sub>O<sub>5</sub> aglycone fragment as the dominant ion (Supplemental Figure 7B in the Supporting Information), analogous to the C<sub>21</sub>H<sub>22</sub>O<sub>10</sub> NG/CNG-hexose isomers described above. The C<sub>13</sub>H<sub>13</sub>O<sub>9</sub> and C<sub>17</sub>H<sub>13</sub>O<sub>6</sub> fragment ions, observed for some NG/CNG-hexosides, also appeared in some of the C<sub>27</sub>H<sub>32</sub>O<sub>15</sub> MS<sup>3</sup> spectra. In addition, the C<sub>23</sub>H<sub>23</sub>O<sub>11</sub> (m/z 475) fragment ion, observed in MS<sup>2</sup> spectra of 4 NG/CNG-dihexosides, was fragmented and the resulting MS<sup>3</sup> spectra were more discriminative than the MS<sup>3</sup> spectra of the C<sub>21</sub>H<sub>21</sub>O<sub>10</sub> fragment ion (Supplemental Figure 7C in the Supporting Information).



**Supplemental Figure 8:** Pie diagrams of the metabolite identification levels of 47 metabolites before and after applying LC-MS<sup>n</sup> to the concentrated tomato extract. The colours represent the different Metabolomics Standards Initiative (MSI) metabolite identification levels, starting on top and going clockwise: 1 (gray), 2 (light gray), 3 (mid gray), and 4 (dark gray). The 47 tomato metabolites were partly characterized or annotated in the Kazusa database (see Supplemental Table 1).

*Supplemental text 1: Experimental details*

*Plant material:* *Lycopersicon esculentum* var. *Gardener's Delight* (a cherry type of tomato fruit) was seeded in January 2008 and grown under standard greenhouse conditions in the Unifarm greenhouse, Wageningen, The Netherlands. Ripe fruits were harvested in May 2008. Directly after harvest, the tomato fruits were peeled and the peel was quickly frozen in liquid nitrogen and mixed to a homogeneous frozen powder. For the Arabidopsis leaf material, a large collection of 350 natural accessions of *Arabidopsis thaliana*, representing a large part of the natural variation present within the species was grown in a climate controlled growth chamber for 29 days under long day length conditions (16 hrs light, 20/18°C) in spring 2010. Each accession was grown in three replicates. Upon harvesting leaves were snap-frozen in liquid nitrogen, pooled and ground to create a homogeneous representative sample. Both plant powders were stored at -80 °C and analyzed by LC-MS<sup>n</sup> within 12 months.

*Metabolite extraction:* For the online LC-MS<sup>n</sup> fragmentation experiments, tomato peel and Arabidopsis leaf were freeze-dried for 40 hours. Then, 150 mg DW was weighed and extracted with 1.5 ml of 75% MeOH in water containing 0.1% formic acid (FA). The sample was vortexed, sonicated for 20 minutes, centrifuged for 15 minutes at 2500 rpm, filtered through a 0.2 µm inorganic membrane filter (Minisart RC4, diameter 4 mm, pore size 0.2 µm; Sartorius), and finally transferred into an HPLC vial. For the LC-fractionation experiments, 500 mg FW material was extracted with 1 ml of 99.9% MeOH + 0.1% FA, following the same procedure as above.

*Details of HPLC-PDA-ESI-MS<sup>n</sup> set up:* The system consisted of an Accela HPLC system with an Accela photodiode array (PDA) detector and autosampler, connected to a LTQ/Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific) that was equipped with an ESI source. Sample injection volume was 5 µl and chromatographic separation took place on a Luna C18(2) analytical column, using a binary eluent solvent system of degassed ultrapure water and acetonitrile, both containing 0.1 % v/v FA (solvent A and B, respectively), at a flow rate of



0.19 mL/min and column temperature at 40°C. The HPLC gradient started at 5% B and linearly increased to 35% B in 45 min. In addition, 15 minutes of washing and equilibration preceded each next injection.

*Full scan and MS<sup>n</sup> settings:* FTMS full scans (m/z 100–1500) were recorded with a resolution of 60,000 (at m/z 400), and with a full AGC target of 200,000 charges, using 2 microscans, whereas for MS<sup>n</sup> scans a resolution of 15,000 (at m/z 400) was used. The capillary temperature was 275 °C, sheath gas flow 60 a.u., auxiliary gas flow 5 a.u., and the sweep gas flow 5 a.u.. The Orbitrap was externally calibrated in negative mode using sodium formate clusters in the range m/z 150-1200 and automatic tuning was performed on m/z 384.93. The Ion trap settings used were: AGC target value of 30,000 and 10,000 in full scan and MS<sup>n</sup> mode respectively. MS<sup>n</sup> settings were programmed as follows: Orbitrap read-out: for all dependent scans, the threshold for further fragmentation was set at 50,000 counts/s, the isolation width was 3 Da, the normalized collision energy was 35%, the activation time was 30,000 ms, and 3 microscans were used for all MS<sup>n</sup> levels. The maximum ion filling times for the Ion trap and FTMS were 100 ms in MS<sup>n</sup> mode and in full scan mode for FTMS 500 ms. Additional AGC targets were 10,000, 20,000, and 30,000 charges for MS<sup>2</sup>, MS<sup>3</sup>, and MS<sup>4</sup> respectively. Ion trap read-out: identical to the Orbitrap settings, but a threshold of 4,500 counts/s and 1 microscan per spectrum were used. If applied, dynamic exclusion settings were set as follows: repeat 1, repeat duration 20 seconds, exclusion list of maximal 500 m/z values, and an exclusion duration of 20 seconds.

*Description of the MS<sup>n</sup> spectral tree topology and online generation:* A small online tree consists of a MS<sup>2</sup> scan and three MS<sup>3</sup> scans of the three most intense fragments ions in MS<sup>2</sup>. A larger online tree comprises two additional MS<sup>4</sup> scans of the two most intense fragment ions in MS<sup>3</sup> and was used during Ion trap read-out. The full scan was always obtained in FTMS read-out. In total 4 different online MS<sup>n</sup> methods were applied, namely the combinations of read-out in the Orbitrap or the Ion trap and with or without dynamic exclusion.

*LC-fractionation of tomato extract:* The system for online fragmentation was adapted with a chip-based nano-electrospray ionization source (Triversa Nano-

Mate, Advion BioSciences). Sample injection volume was 50  $\mu\text{l}$  and chromatographic separation took place on a Luna C18(2) analytical column (Phenomex, 2.0 x 150 mm i.d., 100  $\text{\AA}$ , particle size 3  $\mu\text{m}$ ) with pre-column, using a binary eluent solvent system of degassed ultrapure water and acetonitrile, both containing 0.1 % v/v formic acid (respectively solvent A and B), at a flow rate of 0.19 mL/min. The HPLC gradient started at 5% B and linearly increased to 35% B in 45 min. In addition, 15 minutes of washing and equilibrating preceded each next injection. The flow rate was set at 190  $\mu\text{l}/\text{min}$ . Isopropanol (30  $\mu\text{l}/\text{min}$ ) was added between PDA and NanoMate via a T-junction into the LC flow to ensure a stable nanospray. The NanoMate source was operated in the negative ionization mode with a HD\_A\_384 chip with a spray voltage of 1.7 kV. The NanoMate was used in the LC coupling mode with fraction collection. The total flow (220  $\mu\text{l}/\text{min}$ ) was split using capillary tubes for MS spray (480 nL/min) and for fractionation (219.5  $\mu\text{l}/\text{min}$ ). Spray sensing was used to prevent blockage of the nozzle or charge leakage: as soon as the spray current dropped for more than 90 seconds below 4 nA, or went above 400 nA, the NanoMate automatically continued to the next nozzle, with a maximum of 2 per injection. Fractions were collected by the NanoMate in 5 sec fractions (18  $\mu\text{l}$ ) in a 384 well plate (twin tec, Eppendorf), cooled at 10  $^{\circ}\text{C}$ . After collection, 4  $\mu\text{l}$  isopropanol was added for spray stability of the fractions and the plate was sealed to prevent evaporation of the fractions. Parallel to the fractionation, the MS spray from the NanoMate was used to record FTMS full scans ( $m/z$  100–1500) with resolution of 60.000, and with a full AGC target of 200.000 charges, using 2 microscans. The capillary temperature was 200  $^{\circ}\text{C}$ , sheath gas flow 5 a.u., auxiliary gas flow and the sweep gas flow 0 a.u. The Orbitrap was externally calibrated in negative mode using sodium formate clusters in the range  $m/z$  150-1200 and tuned on  $m/z$  384.9353 and for offline positive mode  $\text{MS}^n$  measurements calibrated using sodium formate clusters in the range  $m/z$  150-1200 and tuned on  $m/z$  362.9263.

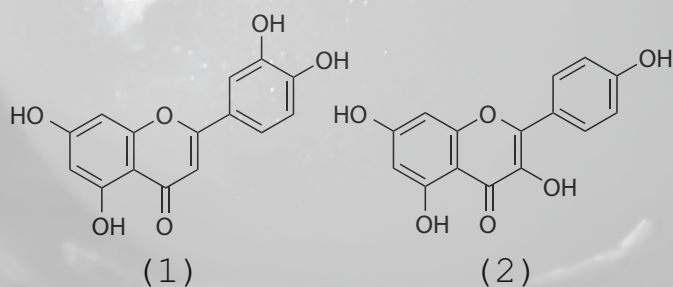
Offline  $\text{MS}^n$  spectral tree generation: Detailed NanoMate settings: in negative mode 0.50 psi back pressure and 1.45 kV spray voltage, and in positive ionization mode 0.40 psi back pressure and 1.40 kV spray voltage was used. The MS1

mass window was reduced from 10 Da to smaller values in case co-eluting higher abundant metabolites were taken as parent ion instead of the selected metabolite.

***Supporting Table 1:*** *This Excel file provides spectral data of the annotated metabolites in this study. The Excel file is available at <http://edepot.wur.nl/216854> and a direct link is: <http://edepot.wur.nl/216857>*



*Chapter 4: A strategy for fast structural elucidation of metabolites in small volume plant extracts using automated MS-guided LC-MS-SPE-NMR*



Justin J.J. van der Hooft, Velitchka Mihaleva, Ric C.H. de Vos,  
Raoul J. Bino, and Jacques Vervoort

This chapter was published as research article in a special issue of Magnetic Resonance in Chemistry, 2011, Volume 49, Issue Supplement S1, pages S55–S60, DOI: 10.1002/mrc.2833

### *Abstract*

In many metabolomics studies, metabolite identification by mass spectrometry (MS) is often hampered by the lack of good reference compounds; and hence, nuclear magnetic resonance spectroscopy (NMR) information is essential for structural elucidation, especially for the very large group of (plant) secondary metabolites. The classical approach for compound identification is to perform time-consuming and laborious high pressure liquid chromatography (HPLC) fractionations and purifications, before (re)dissolving the molecules in deuterated solvents for NMR measurements. Hence, a more direct and easy purification protocol would save time and efforts. Here, we propose an automated MS-guided HPLC-MS-solid phase extraction (SPE)-NMR approach, which was used to fully characterize flavonoid structures present in crude tomato plant extracts. NMR spectra of plant metabolites, automatically trapped and purified from LC-MS traces, were successfully obtained, leading to the structural elucidation of the metabolites. The MS-based trapping enabled a direct link between the mass signals and NMR peaks derived from the selected LC-MS peaks, thereby decreasing the time needed for elucidation of the metabolite structures. In addition, automated  $^1\text{H}$ -NMR spectrum fitting further speeded up the candidate rejection process. Our approach facilitates the more rapid unraveling of yet unknown metabolite structures and can therefore make untargeted metabolomics approaches more powerful.

Keywords: NMR,  $^1\text{H}$ ,  $^{13}\text{C}$ , automation, HPLC-MS-SPE-NMR, metabolomics, polyphenols, structural elucidation, tomato

## *Introduction*

In mass spectrometry (MS)-based metabolomics studies, the unambiguous structural elucidation of molecules is a major bottleneck. MS spectral data usually provide information on the elemental formulas, and MS fragmentation can add valuable structural information,<sup>1</sup> for example in case of conjugated molecules.<sup>2</sup> In case reference compounds are available, MS can give unequivocal identification. However, for structural elucidation of partially or completely unknown metabolites, nuclear magnetic resonance spectroscopy (NMR) is essential.<sup>3</sup> Although NMR is lacking in mass sensitivity compared to MS, NMR provides a much more comprehensive spectroscopic overview of metabolites, enabling structural elucidation by comparing NMR spectra to reference data and analysis of the chemical shifts and fine structure of the observed NMR peaks. For example, NMR determines the exact identity of sugar residues, whereas MS usually only detects the presence of sugar moieties.

Combining spectral data from MS and NMR can considerably decrease the time for structural elucidation of metabolites by candidate rejection and (sub)structure conformation. Identification of a compound is often very difficult based on NMR spectral data only; hence, MS data are also essential. Therefore, high pressure liquid chromatography (HPLC)-based hyphenation of NMR and MS detectors in order to combine sets of spectral information of metabolites has already been used for several years.<sup>4,6</sup> The most hampering factor in the efficient combination of HPLC and NMR was the need for deuterated solvents for NMR experiments of metabolites at relatively low concentrations.<sup>5,7</sup> Solid phase extraction (SPE) was introduced as bridge between HPLC and NMR, enabling evaporation of the HPLC solvents before transferring the analytes with deuterated solvents to the NMR probe.<sup>8,9</sup> In this way, flavonoids and their glycosides were successfully trapped based on their UV signal.<sup>8</sup> Hitherto, signals detected in a LC-Diode Array Detector (DAD) trace were selected for trapping on SPE cartridges. This works well for compounds that contain one or more chromophores, i.e., that are visible through a Photodiode Array (PDA) detector. The multiple trapping of the same peak in different HPLC-runs on the same cartridge concentrated the

analyte, thereby reducing the sensitivity gap between MS and NMR.<sup>5</sup> LC-SPE-NMR in the structural elucidation of metabolites has been used many times in natural product research or in pharmaceutical research.<sup>10, 11</sup> Surprisingly, LC-SPE-NMR has not been reported before for tomato extracts. Moreover, only few compounds from tomato have been characterized by NMR.<sup>12, 13</sup> In this study, peel extracts of tomatoes with modified flavonoid biosynthesis pathways<sup>14, 15</sup> were selected for analysis of the accumulated polyphenols in the tomato fruit peels. Enzymes involved in the formation of flavonoids from malonyl coenzyme A (malonyl-CoA) and phenylpropanoids, for example, chalcone synthase and isomerase, were overexpressed in the tomato mutants, leading to accumulation of mainly quercetin and kaempferol derived flavonoid conjugates in the tomato peel. Flavonoids are an important, well-studied class of polyphenol compounds playing different roles in plant physiological processes,<sup>16, 17</sup> and they are often associated with health benefits for humans.<sup>18</sup> Flavonoids are well studied by MS<sup>19, 20</sup> and NMR,<sup>21</sup> but the vast amount (6000) of structurally closely related molecules<sup>22</sup> pose challenges on the unambiguous elucidation of their structures.<sup>3</sup> To enhance the metabolite identification process, we set up an automated MS-guided HPLC-MS-SPE-NMR method that is able to obtain spectral data for structural elucidation of metabolites like the accumulated polyphenols directly from complex biological extracts. Combined with an automated fitting procedure of the experimental <sup>1</sup>H-NMR spectra, the process of identification of the molecules under study can be remarkably enhanced.

## *Materials and Methods*

### *Materials.*

*Tomato fruit extracts:* The tomato fruit of the genetically modified tomato plants (in a background line of MoneyMaker cultivar)<sup>14, 15</sup> was grown in a greenhouse in Wageningen, The Netherlands, harvested in June 2003, and pooled per plant, after which small pieces of tomato fruit or the frozen powder were stored at -80°C.



*Chemicals:* Acetonitrile (HPLC grade) was purchased from VWR, Leuven, Belgium, acetonitrile NMR chromasolve (for LC-NMR; SPE device) from Riedel-de Haën, Seelze, Germany, and formic acid (99% pure) from Biosolve B.V., Valkenswaard, The Netherlands. Deuterated methanol (99.8% pure) was obtained from CIL Inc., Andover, MA. Ultrapure water was obtained from an in-house Millipore Milli-Q water purification system (A10 gradient, Millipak 40, 0.22 micrometer).

### *Methods.*

*Extraction procedure:* Ten grams of powder fresh weight tomato peel was freeze-dried for 40 hours. Then, 150 mg DW material was extracted with 1.5 mL of 75% methanol (MeOH) in water with 0.1% formic acid. The sample was shortly vortexed and sonicated for 20 minutes, centrifuged for 15 minutes at 2500 rpm, then filtered through a 0.2- $\mu$ m inorganic membrane filter, and finally transferred into an HPLC vial.

*HPLC-MS<sup>n</sup> experiments:* The tomato extracts were injected (5  $\mu$ L) in an LTQ-Orbitrap system (Thermo) using a similar chromatographic set up as used by Moco et al.<sup>23</sup> and fractionated using a NanoMate robot (Advion). Subsequently, selected fractions were fragmented using the Orbitrap.<sup>2</sup> The acquired MS<sup>n</sup> data were analyzed using Xcalibur software.

*HPLC-MS-SPE-NMR experiments:* Chromatographic separation was conducted on an Alltima HP column (Alltima, 4.6 x 150 mm i.d., particle size 3  $\mu$ m) with a pre-column. The same solvents and gradients as for the HPLC-MS<sup>n</sup> experiments were used, at a flow rate of 1 mL/min.

An injection volume of 20  $\mu$ L of each selected sample was subjected to the HPLC-MS-SPE-NMR system that consisted of the online-coupled Agilent 1200 quaternary solvent delivery pump, Agilent 1200 degasser, Agilent 1200 autosampler, Bruker Daltonics MicrOTOF ESI mass spectrometer, Knauer K120 pump for post column water delivery, Spark Prospekt 2 SPE device containing Oasis (Waters Company) HLB cartridges (2.0  $\times$  10 mm i.d.), and the Bruker Advance III 600 spectrometer.

To ensure reproducibility of the chromatogram before the peak trapping experiments, two pre-injections were done. When a stable reproducible chromatogram was obtained, the second chromatogram was loaded in Hystar to determine thresholds based on the base peak chromatogram. The SPE cartridges were washed with 500  $\mu\text{L}$  acetonitrile and equilibrated with water containing 0.1% v/v formic acid (solvent A), in order to be ready for peak trapping.

A Bruker splitter (5:95) was mounted on top of the ESI source and used to direct 5% of the eluent flow to the MicrOTOF (50  $\mu\text{L}/\text{min}$ ) and 95% towards the SPE device. The Knauer K120 pump served as make-up pump and added solvent A (flow rate 1 ml/min) directly after the splitter. The SPE device was used to automatically trap the selected peaks based on thresholds (counts per second) set in the base peak chromatogram (BPC), detected by the MicrOTOF. Three cumulative trappings in multi trapping mode were performed, before the cartridges were put to dry 59 minutes (the maximum setting in the software) with nitrogen gas.

The following MicrOTOF settings were used at a flow rate of 50  $\mu\text{L}/\text{min}$ : capillary voltage +3.2 KV, nebuliser pressure 0.8 bar, dry gas flow 7.0 L/min, dry temperature 180 °C, capillary exit -150.0 V, hexapole RF 150.0 Vpp, lens 1 transfer time 63  $\mu\text{s}$ , and the lens 1 pre pulse storage time 1  $\mu\text{s}$ , and a mass window of  $m/z$  100 – 1500. The MS method consisted of three segments, generating a Na-Formate peak in each chromatogram by valve switching between 0.2 and 0.5 minutes retention time. External calibration took place before each series of measurements with a series of Na-Formate adducts in the  $m/z$  100 – 1200 range. In addition, internal calibration was possible by using the calibrant (sodium formate adducts) peak present in each chromatogram.

Finally, the content of each cartridge was flushed with deuterated acetonitrile (with a pre-defined instrument specific volume of 227  $\mu\text{L}$ ) via capillaries to a Bruker Avance III 600 spectrometer equipped with a 5 mm inverse cryogenic flow probe, with the Bruker cryofit (active volume 30  $\mu\text{L}$ ) inserted, operating at 300 K. After NMR measurements, the probe was flushed and the content was collected in a HPLC vial and stored at 4 °C. Of each metabolite, 8  $\mu\text{L}$  was used

for MS<sup>n</sup> experiments as described in van der Hooft et al.<sup>2</sup>

For each cartridge, 1D-<sup>1</sup>H-NMR measurements were performed at 600 MHz using standard pulse sequences. If the purity and signal-to-noise ratio of the obtained 1D-<sup>1</sup>H-NMR spectra were sufficient, two-dimensional (2D)-<sup>1</sup>H and heteronuclear measurements were conducted. The first transferred sample was manually shimmed, after which TOPSHIM was used to automatically shim the other transferred samples in one measurement series. The spectra were automatically phase- and baseline corrected.

Hystar Compass 1.3 controlled the data acquisition, as well as peak trapping and transferring to the NMR spectrometer. Bruker Topspin 2.1 was used for NMR data acquisition and processing.

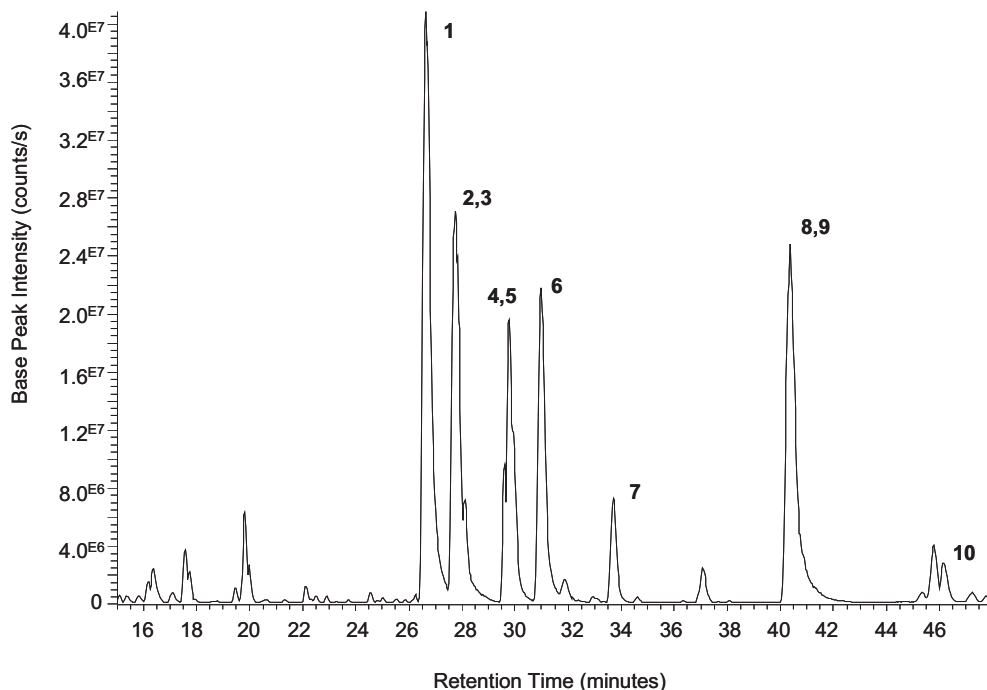
*Structure elucidation:* The automated consistency analysis module<sup>24</sup> from the PERCH NMR software<sup>25</sup> was used to fit the experimental spectra using predicted spectral parameters of a set of compounds. This set was extracted by querying a newly developed MetIDB database (to be published) using chemical formula or exact mass obtained from MS data. The set was further refined by comparing the experimental chemical shifts and the predicted values of the database compounds. Within MetIDB over 6000 polyphenolic molecules are stored with reliable three-dimensional structures, monoisotopic masses, and NMR characteristics. The chemical shifts in the library and the query spectra were binned with a window size of 0.1 ppm. The score of the compounds was defined as the sum of the absolute differences between the bins of the query and the database compounds. In the final set only compounds with scores smaller than number of query shifts divided by 2 were retained. For example, if the query contains six chemical shifts and for two chemical shifts no match was found in MetIDB, a score of four is assigned to the database compound. This database compound is excluded from the list as  $4 > (6/2)$ . Automated spectrum fit requires clean samples and high signal to noise ratio. As the compounds were concentrated on a SPE cartridge their spectra always contained signals from formic acid and acetonitrile. The regions in which these signals appear needed to be excluded from the analysis.

## Results

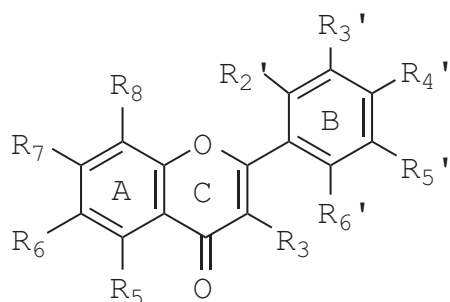
Crude extracts of tomato lines altered in the flavonoid biosynthesis pathway were subjected to reversed phase C18 chromatography prior to automated MS-guided SPE-NMR trapping. The eluent flow was splitted such that a small part (5%) was infused into the MicrOTOF MS, and that the major part (95%) went through the Prospekt SPE device together with the make-up flow. Trapping of the analytes was based on the intensity of the LC-MS peaks, such that when the intensity of the base peak was above a certain threshold, the flow of the column was redirected toward the HLB SPE cartridges. Trapping of the analytes could thus be performed multiple times in a row, with each detected analyte trapped on a specific cartridge. After drying, the analytes were transferred with deuterated methanol to the NMR spectrometer, with subsequent 1D and 2D NMR data acquisition. After the NMR measurements the analytes could be flushed to a vial for subsequent MS<sup>n</sup> analysis on a LTQ-Orbitrap system<sup>2</sup>.

In Figure 1 the LC-MS chromatogram of a tomato extract is shown. A large number of components is visible in the MS spectrum. The components observed in the chromatogram were automatically trapped three times. The MicrOTOF MS data and the accurate mass MS and MS<sup>n</sup> spectral data were used for elemental formula calculations.

A database of predicted one dimensional (1D)-<sup>1</sup>H-NMR chemical shifts can be of great help for reducing the number of candidate compounds. This is illustrated with the following two examples. A chemical formula of C<sub>15</sub>H<sub>10</sub>O<sub>7</sub> was assigned to peak 8 in Figure 1. The MetIDB database contains 22 isomers all corresponding to compounds with this chemical composition. This list of 22 isomers was reduced to 4 compounds by filtering on a match for the <sup>1</sup>H-NMR chemical shifts observed to the predicted <sup>1</sup>H-NMR chemical shifts. A window size of 0.1 ppm was used for this filtering. The positions of the hydroxyl groups and hydrogen atoms in the four compounds are listed in Table 1. In these 4 molecules the hydroxyl groups are present on positions 5 and 7 in the A-ring (see Figure 2). In hits 2 and 4 the B-ring is *para* substituted. Due to the symmetry of the B-ring of hit 2 and 4, the predicted spectrum cannot be fitted onto the experimental



**Figure 1:** Part of LC-MS chromatogram of a flavone/flavonol biosynthesis tomato mutant extract. The peaks selected for automated MS-guided SPE-NMR trapping are indicated.



**Figure 2:** Structure and labeling of the position of the substituents of flavones.

spectrum as the intensity of signals and the coupling patterns are different. Hit 1 and 3 have the same spin particle systems and the fit was successful for both compounds.

However, the correlation coefficient and the standard deviation for hit 1 is much better than for hit 3. Hence, the compound corresponding to peak 8 was identified as luteolin. In the second example, peaks 2 and 3 were trapped in the same cartridge with peak 2 being the major component. A set of 46 compounds was selected in MetIDB by querying for the chemical formula of peak 2 being  $C_{21}H_{20}O_{12}$ . This set was further reduced to 7 compounds using the set of experimentally determined  $^1H$ -NMR values. MS/MS data were

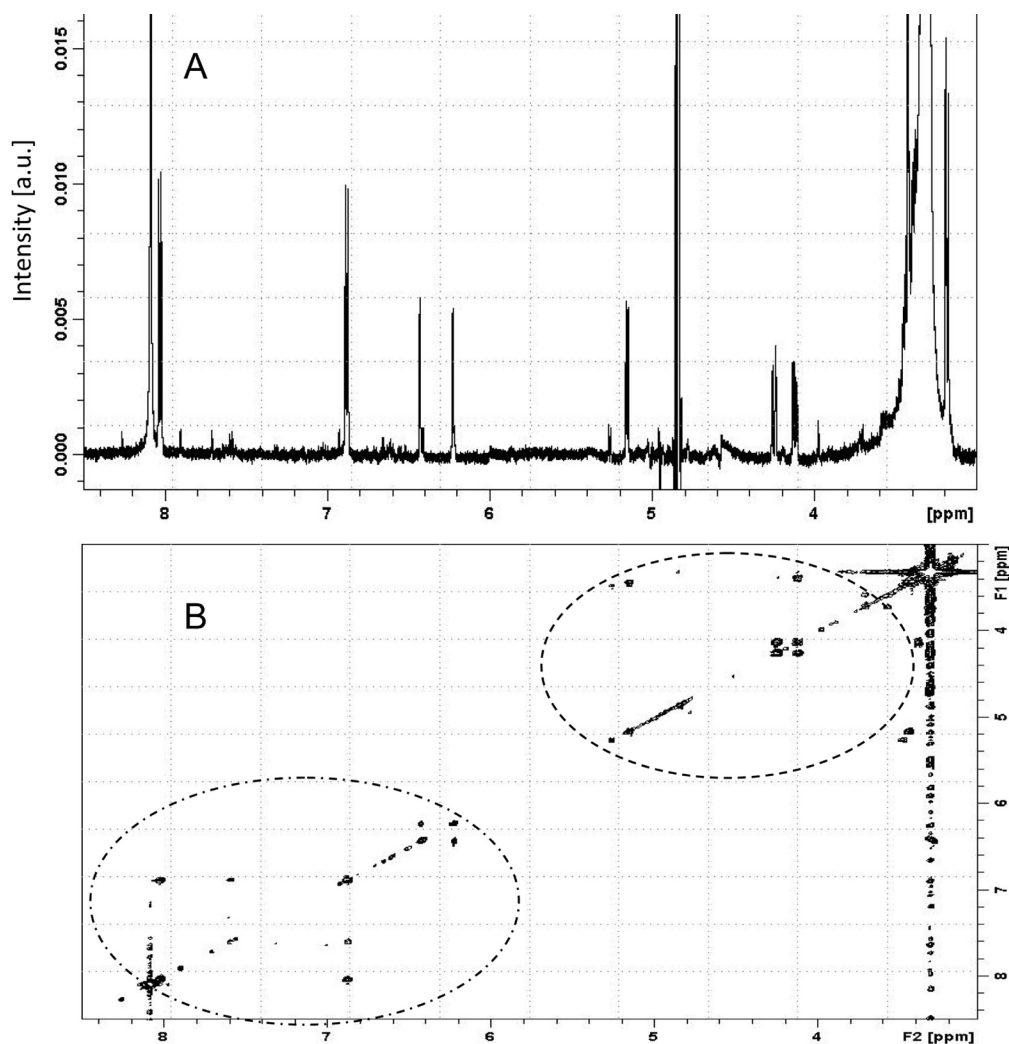
**Table 1:** Correlation coefficient (*R*) and the standard deviation (*SD*) between predicted and fitted chemical shifts of the aromatic protons of positional isomers of compounds 2 and 8. The proton, hydroxyl, and sugar moiety positions correspond to the labels indicated in Figure 2.

Library hit list	OH positions	H positions	Sugar	R	STD
<b>Peak 8</b>					
1	R <sub>3'</sub> , R <sub>4'</sub> , R <sub>5</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>3</sub> , R <sub>6</sub> , R <sub>8</sub>	-	1.00	0.04
2	R <sub>4'</sub> , R <sub>5</sub> , R <sub>7</sub> , R <sub>8</sub>	R <sub>2'</sub> , R <sub>3'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>3</sub> , R <sub>6</sub>	-	ND	ND <sup>a</sup>
3	R <sub>3'</sub> , R <sub>6'</sub> , R <sub>5</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>4'</sub> , R <sub>5'</sub> , R <sub>3</sub> , R <sub>6</sub> , R <sub>8</sub>	-	0.90	0.20
4	R <sub>4'</sub> , R <sub>5</sub> , R <sub>6</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>3'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>3</sub> , R <sub>8</sub>	-	ND	ND
<b>Peak 2</b>					
1	R <sub>3'</sub> , R <sub>4'</sub> , R <sub>5</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>6</sub> , R <sub>8</sub>	R <sub>3</sub> , O-β-glu	1.00	0.01
2	R <sub>3'</sub> , R <sub>4'</sub> , R <sub>5</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>6</sub> , R <sub>8</sub>	R <sub>3</sub> , O-β-gal	1.00	0.01
3	R <sub>3'</sub> , R <sub>4'</sub> , R <sub>3</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>6</sub> , R <sub>8</sub>	R <sub>5</sub> , O-β-glu	1.00	0.06
4	R <sub>3'</sub> , R <sub>4'</sub> , R <sub>5</sub> , R <sub>7</sub>	R <sub>2'</sub> , R <sub>5'</sub> , R <sub>6'</sub> , R <sub>3</sub> , R <sub>6</sub>	R <sub>8</sub> , O-β-glu	0.91	0.25

<sup>a</sup>ND – R and STD were not defined for this compound.

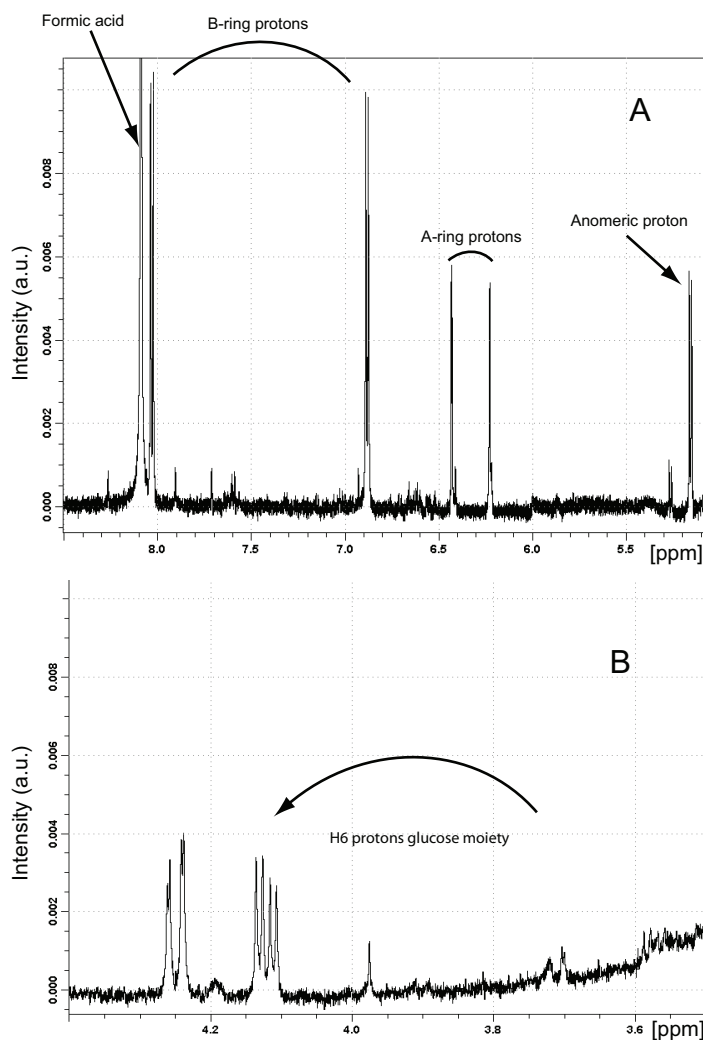
used to reject two compounds having aurone structure and another compound with a C-linked glucose. The remaining four compounds are listed in Table 1. Due to the presence of small amounts of compound 3 in the spectra and an overlap of one of the sugar protons with the methanol signal, an automated fit of the predicted spectrum to the complete 1D-<sup>1</sup>H-NMR spectrum was not possible. Therefore, only the aromatic region was used for automated analysis. Hits 1 and 2 differ in the sugar moiety. Compared to hit 1, in hit 3 the glucose fragment is attached to the A-ring at the 5th position. Hit 4 showed the largest deviation from the experiment and was thus excluded from further analysis. Glucose and galactose can be distinguished due to differences in the splitting pattern of the proton attached to C4 atom. A manual PERCH fit of the 5.5-3.0 ppm region was successful with glucose but not with galactose. The MS<sup>n</sup> spectral tree of peak 2 showed the attachment of a sugar to the C3 position,<sup>2</sup> leading to the elucidation of compound 2 as quercetin-3-*O*-glucoside.

Because of low signal-to-noise ratio and overlap of the sugar protons with



**Figure 3:** 1D- $^1\text{H}$ -NMR spectrum (A) and 2D-COSY spectrum (B) of peak 7. The three intense peaks in A at 8.0, 4.9, and 3.3 ppm belong to formic acid, residual  $\text{H}_2\text{O-D}$ , respectively  $\text{MeOH-d}_3$ .

the solvent signal, it is not always possible to resolve all sugar protons. The anomeric and the H6 protons of a sugar moiety are usually well resolved as, for example, in the NMR spectra obtained from peak 7 (shown in Figure 3). The mass of peak 7 was observed to be  $m/z$  533.0937  $[\text{M-H}]^-$  to which the elemental formula  $\text{C}_{24}\text{H}_{21}\text{O}_{14}$  was assigned. The in-source loss of  $\text{CO}_2$  was clearly observed, which is indicative for the presence of a malonyl group, resulting in a mass of  $m/z$  489.1040  $[\text{M-H}]^-$  ( $\text{C}_{23}\text{H}_{21}\text{O}_{12}$ ). In the 1D- $^1\text{H}$ -NMR



**Figure 4:** (A) Enlarged part of the  $1D-^1H$ -NMR spectrum of peak 7 with diagnostic protons of the kaempferol unit indicated. The B ring refers to the para-substituted aromatic ring, whereas the A ring refers to the dihydroxy-substituted ring of compound 7. (B) The H6 protons of the glucose moiety were shifted significantly compared to their usual shifts around 3.72 and 3.54 ppm.

spectrum, typical signals for kaempferol were visible, and in addition it was observed that the sugar H6 protons observed for compound 7 were shifted by 0.5 ppm (see Figure 4) compared to kaempferol-3-*O*-glucoside (compound 6), which is an indication for the presence of electron withdrawing group in their neighborhood. This indicates that the malonyl group is attached to the glucose-6 position. This observation is also independently supported by the fact that only one compound having the structure of kaempferol-3-*O*-(6-*O*-malonyl)glucoside can be retrieved from MetIDB using the aromatic, anomeric, and the H6 proton chemical shifts as input variables for the query. Thus, the presence of a malonyl



**Table 2:** Identified compounds as observed in Figure 1.

Peak #	Compound name	OH positions	H positions	Sugar <sup>a</sup>
1	Quercetin-3- <i>O</i> -rutinoside	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -rutinoside
2	Quercetin-3- <i>O</i> -glucoside	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> , R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -β-glu
3	Luteolin-7- <i>O</i> -glucoside	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>3</sub> ,R <sub>6</sub> , R <sub>8</sub>	R <sub>7</sub> , 7- <i>O</i> -β-glu
4	Kaempferol-3- <i>O</i> -rutinoside	R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>3</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -β-rutinoside
5	Quercetin-3- <i>O</i> -(6- <i>O</i> -malonylglucoside)	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> , R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -β-glu
6	Kaempferol-3- <i>O</i> -glucoside	R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>3</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -β-glu
7	Kaempferol-3- <i>O</i> -(6- <i>O</i> -malonylglucoside)	R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>3</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>6</sub> ,R <sub>8</sub>	R <sub>3</sub> , 3- <i>O</i> -β-glu-6- <i>O</i> -malonyl
8	Luteolin	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> , R <sub>3</sub> ,R <sub>6</sub> ,R <sub>8</sub>	-
9	Quercetin	R <sub>3</sub> <sup>'</sup> ,R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub> , R <sub>3</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>6</sub> ,R <sub>8</sub>	-
10	Apigenin	R <sub>4</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>7</sub>	R <sub>2</sub> <sup>'</sup> ,R <sub>3</sub> <sup>'</sup> ,R <sub>5</sub> <sup>'</sup> ,R <sub>6</sub> <sup>'</sup> ,R <sub>3</sub> ,R <sub>6</sub> ,R <sub>8</sub>	-

<sup>a</sup>Glucose and galactose are denoted as glu and gal, respectively

group was supported both by <sup>1</sup>H-NMR and MS data and peak 7 was elucidated as kaempferol-3-*O*-(6-*O*-malonylglucoside). The complete list of the identified compounds is listed in Table 2.

## Discussion

LC-SPE-NMR was established less than a decade ago and was shown to be very useful for pharmaceutical and/or natural product research. The major advantage of LC-SPE-NMR is that compounds of interest can be concentrated into a small volume, which, especially with the use of small-volume cryogenically cooled NMR radiofrequency (rf) coils enables the acquisition of NMR spectra with excellent S/N. Selection for the trapping of compounds of interest was based on UV absorbance of molecules separated on a column. This selection criterion based

on UV absorbance caused several problems. Firstly, many molecules do not have a chromophore. Secondly, molecules that were overlapping in a chromatographic run and that could not be distinguished based on their UV absorbance, yielded spectra with overlapping NMR signals. In theory, by applying diffusional filters, smaller molecules can be separated from larger molecules in the NMR spectra, but this is not straightforward to do; moreover, it is a time-consuming NMR procedure.<sup>3,26</sup> Because a mass spectrometer is an excellent detector also for non-chromophores (for instance, alkaloids), we decided in our new approach to set up an automated trapping system not based on the UV signal but based on the signal of the mass spectrometer. Moreover, the mass signal in most cases enables the assignment of an elemental formula to the metabolite, thereby providing already important information for candidate selection. The set up for automated trapping based on MS is more advanced than trapping based on UV as the chromatography needs to be very stable, due to the use of additional capillaries and splitters. The use of in-house build splitters is an interesting alternative to the industrially made splitters in order to stabilize the chromatography in case of lower intense or less well-defined peaks and direct 1-2% of the flow towards the MS instead of 5%.

An important aspect of the whole automated MS-guided trapping procedure is the trapping efficiency of the cartridges. There are many different sorbent materials for the SPE cartridges and from the sorbents we tested HLB (Oasis), which is a relatively new reversed phase sorbent, turned out to work very efficiently for trapping polyphenolic molecules. The HLB sorbent of Oasis is used in the SPE cleanup of samples; however, the use of HLB sorbent in cartridges for SPE-NMR is still not wide-spread.<sup>27, 28</sup>

The tomato extracts used for this study were known to contain accumulated levels of flavonols and flavones, because of overexpression of genes in the flavonoid biosynthesis pathway.<sup>14, 15</sup> In the study of Schijlen et al.,<sup>15</sup> only a limited number of flavonoids could be identified based on LC-MS data for the selected retention times in the chromatogram. Therefore, we decided to use these samples for automated MS-guided SPE-NMR trapping. The accumulation of

metabolites in the peel<sup>14, 15</sup> enabled acquisition of good quality 1D-<sup>1</sup>H-NMR and in many cases even 2D-COSY and 2D-HSQC spectra were obtained from only three automated trappings. Identification of the compounds measured was based on the accurate mass obtained, on the NMR signals from 1D and/or 2D datasets, on an accurate fitting of the spectra using the PERCH NMR Software<sup>21, 25</sup> and on the MS<sup>n</sup> fragmentation patterns obtained.<sup>2</sup> In addition, we tested the automated procedure for metabolite identification using the MetIDB database. MS and NMR data obtained from automated MS-guided LC-SPE-NMR were automatically matched against the database providing quick and reliable identifications in an extremely short time for many molecules under study. Currently, about 2 µgram of pure compound is sufficient for an excellent NMR spectrum that can be compared to reference experimental or predicted 1D-<sup>1</sup>H-NMR data.

### *Conclusions*

The MS guided HPLC-MS-SPE-NMR proved to be an efficient tool in the structural elucidation process of metabolites in complex crude plant extracts. The direct coupling between LC-MS and NMR spectral data decreases the time needed for unambiguous elucidation of metabolite structures. Moreover, the use of analytical HPLC separation is preferred over the use of preparational HPLC separation due to better peak separation and the less sample material that is needed. Following our approach, we were able to automatically trap 10 LC-MS peaks and identify the structures of their analytes. The combination of MS and MS<sup>n</sup> spectral data and predicted 1D-<sup>1</sup>H-NMR spectral characteristics proved to be a powerful tool for candidate selection and rejection. With the still increasing sensitivity of NMR and improvements in automated annotation of NMR spectra, we believe that the connection between MS and NMR will become increasingly fruitful.

## Acknowledgements

This research was granted by the Netherlands Metabolomics Centre and the Centre for BioSystems Genomics, both of which are part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

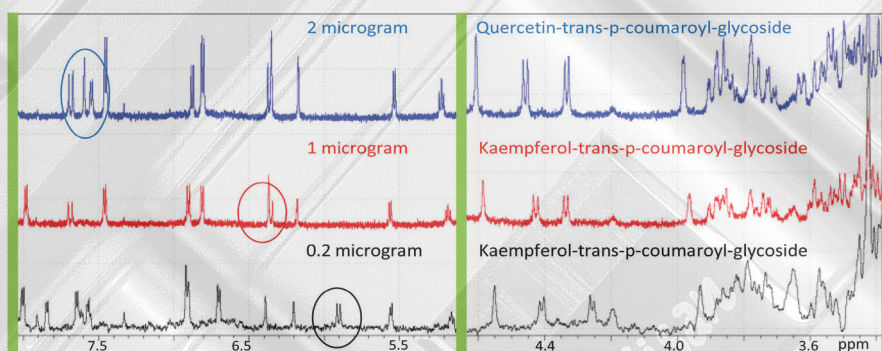
## References

1. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23-60.
2. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino, J. Beekwilder and R. C. H. De Vos, *Anal. Chem.*, 2011, **83**, 409-416.
3. S. Moco, J. Vervoort, R. J. Bino, R. C. H. De Vos and R. Bino, *Tr. Anal. Chem.*, 2007, **26**, 855-866.
4. J. L. Wolfender, C. Terreaux and K. Hostettmann, *Pharmaceut. Biol.*, 2000, **38**, 41-54.
5. J. W. Jaroszewski, *Planta Med.*, 2005, **71**, 795-802.
6. J. W. Jaroszewski, *Planta Med.*, 2005, **71**, 691-700.
7. S. Agnolet, J. W. Jaroszewski, R. Verpoorte and D. Staerk, *Metabolomics*, 2010, **6** **292-302**.
8. C. Clarkson, D. Staerk, S. Honoré Hansen and J. W. Jaroszewski, *Anal. Chem.*, 2005, **77**, 3547-3553.
9. V. Exarchou, M. Godejohann, T. A. v. Beek, I. P. Gerotheranassis and J. Vervoort, *Anal. Chem.*, 2003, **75**, 6288-6294.
10. T. Grevnstuk, J. J. J. van der Hoof, J. Vervoort, P. de Waard and A. Romano, *Biochem. System. Ecol.*, 2009, **37**, 285-289.
11. H. Tang, C. Xia and Y. Wang, *Magn. Res. Chem.*, 2009, **47**, S157-S162.
12. R. Slimestad, T. Fossen and M. J. Verheul, *J. Agr. Food Chem.*, 2008, **56**, 2436-2441.
13. R. Slimestad and M. Verheul, *J. Sci. Food Agricult.*, 2009, **89**, 1255-1270.
14. A. Bovy, E. Schijlen and R. D. Hall, *Metabolomics*, 2007, **3**, 399-412.
15. E. Schijlen, C. H. Ric De Vos, H. Jonker, H. Van Den Broeck, J. Molthoff, A. Van Tunen, S. Martens and A. Bovy, *Plant Biotech. J.*, 2006, **4**, 433-444.
16. J. B. Harborne, *Comparative Biochemistry of the flavonoids* 1967.
17. E. de Rijke, P. Out, W. M. A. Niessen, F. Ariese, C. Gooijer and U. A. T. Brinkman, *J. Chromatogr. A*, 2006, **1112**, 31-63.
18. M. E. Obrenovich, N. G. Nair, A. Beyaz, G. Aliev and V. P. Reddy, *Rejuven. Res.*, 2010, **13**, 631-643.
19. V. Vukics and A. Guttman, *Mass Spectrom. Rev.*, 2010, **29**, 1-12.
20. M. Stobiecki and P. Kachlicki, *Acta Physiolog. Plant.*, 2005, **27**, 109-116.
21. S. Moco, L. H. Tseng, M. Spraul, Z. Chen and J. Vervoort, *Chromatographia*, 2006, **64**, 503-508.
22. M. Arita, <http://www.metabolomics.jp>, 2011.
23. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and C. H. R. de Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.
24. H. Thiele, G. McLeod, M. Niemitz and T. Kühn, *Monatshefte für Chemie*, 2011.
25. PERCH, <http://www.perchsolutions.com>, 2011.

26. R. Novoa-Carballal, E. Fernandez-Megia, C. Jimenez and R. Riguera, *Nat. Prod. Rep.*, 2010.
27. C. Clarkson, M. Sibumb, R. Mensenb and J. W. Jaroszewski, *J. Chromatogr. A*, 2007, **1165**, 1–9.
28. A. Espada, C. Anta, A. Bragado, J. Rodríguez and C. Jiménez, *J. Chromatogr. A*, 2011, **1218**, 1790-1794.



## Chapter 5: Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts



Justin J.J. van der Hooft, Moktar Akermi, Fatma Yelda Ünlü,  
Velitchka Mihaleva, Victoria Gomez Roldan, Raoul J. Bino,  
Ric C. H. de Vos, and Jacques Vervoort

This chapter was published as research article in a special issue of Journal of Agricultural and Food Chemistry, 2012, online article ASAP, DOI: 10.1021/jf300297y



### *Abstract*

In order to obtain more insight into the complex phenolic composition of tea, we applied advanced analytical approaches consisting of both LC-LTQ-Orbitrap Fourier Transformed (FT)-mass spectrometry (MS) and LC-time-of-flight-(TOF)-MS coupled to solid phase extraction (SPE)-nuclear magnetic resonance spectroscopy (NMR). Based on the combined structural information from i) accurate multistage mass fragmentation ( $MS^n$ ) spectra, derived by using LC-Orbitrap FTMS<sup>n</sup>, and ii) proton-NMR spectra, derived after LC-TOFMS triggered SPE trapping of selected compounds, 177 phenolic compounds were annotated. Most of these phenolics were glycosylated and acetylated derivatives of flavan-3-ols and flavonols. Principal component analysis based on the relative abundance of the annotated phenolic compounds in 17 commercially available black, green and white tea products separated the black teas from the green and white teas, with epicatechin-3,5-di-*O*-gallate and prodelphinidin-*O*-gallate being among the main discriminators. The results indicate that the combined use of LC-LTQ-Orbitrap FTMS and LC-TOFMS-SPE-NMR leads to a more comprehensive metabolite description and comparison of tea and other plant samples.

Keywords: Ellagic acid, MS,  $MS^n$ , Metabolite identification, NMR, Phenolic conjugates, Polyphenols, Profiling, Tea.



## *Introduction*

Tea drink is prepared as a hot water infusion from tea material, that is, leaves or buds of the *Camellia sinensis* plant, and is one of the most consumed beverages in the world; its consumption has been positively associated with human health.<sup>1-3</sup> The chemical content of the tea drink and, thereby, its flavor, taste and health characteristics, depends on many factors such as postharvest treatment, genotype, growing conditions and plant processing.<sup>4</sup> White tea is an infusion of the dried young leaf buds, green tea derives from steamed and dried mature green leaves, and black tea is an infusion of the fermented and heated leaves. Polyphenols, polyphenol conjugates, and polymerized phenolic structures are the main constituents of the tea drink. Due to the fermentation process, polymerized phenolic structures such as theaflavins and thearubiginins and larger related phenolic complexes can be found in black tea.<sup>5</sup> In addition, a diverse mixture of conjugated flavonoids is present in black, green, and white teas.<sup>6</sup> The number of glycoside moieties attached to phenolic aglycones such as kaempferol, quercetin, and myricetin ranges from one to four, and also acyl groups like *p*-coumaroyl can be attached, forming complex conjugates.<sup>7</sup> The complete structural identification of these conjugates is important to improve the link between the tea quality and nutritional value and its chemical variation.<sup>8</sup> Hyphenated analytical techniques such as liquid chromatography (LC) coupled to mass spectrometry (MS),<sup>9</sup> have been used to study the chemical content of tea samples.<sup>1, 6, 7, 10-13</sup> Quality assessments and influences of production and processing effects on specific tea metabolites have also been described.<sup>14, 15</sup> In view of their potential health benefits, most of these studies have been directed towards the class of phenolic compounds, which include polyphenols and hydroxycinnamic acids.<sup>16</sup> Because authentic standards are scarcely available, complete structural elucidation is often not achieved by using only LC-MS or LC-MS/MS. Thus, metabolite annotation and identification remain major challenges in tea metabolomics.

In the present research we applied LC coupled to an on-the-fly multistage accurate mass Orbitrap Fourier transformed MS (FTMS<sup>n</sup>) approach in pair with

LC coupled to a time-of-flight (TOF) MS-based solid-phase-extraction (SPE)-nuclear magnetic resonance spectroscopy (NMR) method<sup>17, 18</sup> in order to i) get a more comprehensive picture of the complex metabolome of tea products and to ii) fully identify selected tea metabolites. Here, we focused on the annotation of the main phenolic acids and polyphenols present. Specifically, we studied the complex conjugation patterns of the flavan-3-ols catechin and epicatechin, and the flavonols kaempferol and quercetin. The structural information from accurate mass LC-MS<sup>n</sup>, that is, elemental formulas and fragmentation patterns, was combined with diagnostic signals in the one-dimensional (1D)-proton (<sup>1</sup>H)-NMR spectra for their chemical elucidation, leading to a more comprehensive insight into the metabolome of tea products.

### *Material and Methods*

*Tea samples:* Black, green, and white tea samples were obtained from different origins: a local store (bought on June 5, 2011), and a dedicated tea shop (bought on June 4, 2011) that also ordered several special tea samples at <http://www.wollenhaupt.com/> (article no. in brackets) with packaging date 19-05-2011, which are indicated with an asterisk. The tea products were stored under dark conditions at room temperature until use (about 6 weeks). Tea samples were labeled as follows: green tea (GT) - China Yunnan (Y) [article no. 00518]\*, China Sencha (S) [article no. 00515]\*, Chinese Huangshan Green Tea from lab (GL) (local store), China Bancha (B) [article no. 00568], China Chun Mee (CCM) [article no. 00500]\*, China Lung Ching (CLC), China Gunpowder (CG); white tea (WT) - China Mao Feng (CMF) [article no. 00519]\*, China Pai Mu Tan (PMT) [article no. 00509]\*, Jasmijn and Oranjebloesem (WTJ) (local store); black tea (BT) - Darjeeling FTGFOP1 (D) [article no. 00312]\*, Ceylon OP Adawatte (CAD) [article no. 00203]\*, Assam TGFOP1 Hazelbank (A) [article no. 00109]\*, English Earl Grey – Ceylon Black Tea with Bergamot (CB) [article no. 10861], Irish Breakfast (IB) article no. 00418], Ceylon OP Pettiagalla (CP) [article no. 00209], Pickwick English Tea Blend (BL) (local store).

*Chemicals:* Acetonitrile (HPLC grade) was obtained from Biosolve

(Valkenswaard, The Netherlands), methanol (HPLC grade) from Merck-Schuchardt (Hohenbrunn, Germany), formic acid (99-100) from VWR international SAS (Briare, France), acetonitrile NMR chromasolve (for LC-NMR; SPE device) from Riedel-de Haën, Seelze, Germany, and deuterated methanol (99.8% pure) was purchased from CIL Inc., Andover, MA. Ultrapure water was obtained from an in-house Millipore Milli-Q water purification system (A10 gradient, Millipak 40, 0.22 µm).

*Metabolite extraction:* The methanol/water extraction was performed as follows: a few grams of dry tea was ground into a fine powder, and 100 mg was weighed and extracted with 10.00 mL of methanol-water (60:40, v/v) using a sonicator (40 KHz, 100W) for 60 min at room temperature. The slurry mixture was centrifuged at 2500 rpm for 15 min. The supernatant was filtered through a 0.45 µm PVDF syringe filter prior to the solid phase extraction step.

*Solid Phase Extraction:* HLB SPE cartridges (3 cm<sup>3</sup>, OASIS, Waters) were activated under vacuum with 4 mL of methanol and washed with 6 mL of H<sub>2</sub>O. Then, 1.5 mL of extracted tea sample was diluted 6:1 with H<sub>2</sub>O followed by another 1:1 dilution with H<sub>2</sub>O containing 4% phosphoric acid (final pH about 3.0) and deposited on the SPE cartridges, followed by a washing with 4 mL of H<sub>2</sub>O. Elution of tea compounds was performed with 4 mL of 100% methanol. Subsequently, the eluates were dried overnight in a Speedvac protected from light at 35°C and then stored at -20 °C until further use (within 5 days).

*Sample preparation for analysis:* For LC-MS<sup>n</sup>, the freeze dried samples were redissolved in 200 µL of 75% methanol in H<sub>2</sub>O containing 0.1% formic acid, while for LC-MS-SPE-NMR, the freeze dried samples of eight SPE extractions (thus 12 mL of sample) were combined and redissolved in 800 µL of 75% MeOH in H<sub>2</sub>O containing 0.1% formic acid. Samples were then sonicated for 5 min and filtered through a 0.45 µm filter before transferring to a HPLC vial.

*LC-MS<sup>n</sup> conditions:* The C18-reversed phase LC-MS<sup>n</sup> set up consisted of an Accela HPLC tower connected to a photodiode array (PDA) detector and an LTQ/Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific). LC conditions were as described earlier,<sup>19</sup> using a binary eluent solvent system of degassed

ultrapure water and acetonitrile, both containing 0.1% v/v formic acid (solvent A and B, respectively), with an injection volume of 5  $\mu\text{L}$ , a flow rate of 0.19 mL/min and a column temperature of 40 °C. Eluting compounds were ionized and subsequently trapped within an LTQ Ion trap followed by MS<sup>n</sup> fragmentation with accurate mass read-out. MS settings and spectral tree topologies were as previously described.<sup>18</sup>

*HPLC-MS-SPE-NMR conditions:* The HPLC-MS-SPE-NMR system was used as described in ref<sup>17</sup>. In short, it consisted of the online-coupled Agilent 1200 quaternary solvent delivery pump, Agilent 1200 degasser, Agilent 1200 autosampler, Bruker Daltonics MicrOTOF ESI mass spectrometer, Knauer K120 pump for post column water delivery, Spark Prospekt 2 SPE device containing Oasis (Waters Co.) HLB cartridges (2.0 x 10 mm i.d.), and the Bruker Advance III 600 spectrometer. Chromatographic separation was on an Alltima HP column (Alltima, 4.6 x 150 mm i.d., particle size = 3 mm) with a pre-column of the same material. For black (Pickwick English Tea Blend) and green (China Bancha) teas, 4 and 5, respectively, cumulative trappings of selected compounds were performed, based on their LC-MS base peak, using 40  $\mu\text{L}$  injection volume for each trapping event.

*NMR experiments:* The content of each cartridge was transferred to the NMR spectrometer by eluting with 227  $\mu\text{L}$  MeOD, followed by the acquisition of 1D-<sup>1</sup>H-NMR measurements of 128 scans and a  $d_1$  of 4 s (i.e., almost 15 min) with a receiver gain of 128 in MeOD at 600 MHz NMR (Bruker Daltonics) using standard pulse sequences. Two-dimensional (2D) <sup>1</sup>H measurements (COSY) were conducted for structural confirmation if the purity and signal-to-noise ratio of the obtained 1D-<sup>1</sup>H spectra were sufficient (i.e.,  $\geq 8$ ). Quantification of the complex conjugates in the NMR probe was performed by comparing proton integrals to those obtained from a 3 mm NMR tube filled with 200  $\mu\text{l}$  of a standard flavonoid (rutin) dissolved in MeOD at a concentration of 13  $\mu\text{g/mL}$ .

*Metabolite identification:* Metabolites were identified by querying the assigned elemental formulae in Web-based databases such as Chemspider (<http://www.chemspider.com/>) and Scifinder (<https://scifinder.cas.org/scifinder/>). After the

MS<sup>n</sup> fragmentation patterns had been studied, the chemical structures of the flavonoid backbones were used as restraint. If NMR data was available, further candidate selection was performed by comparing observed NMR spectral data to those in literature. The Dictionary of Natural Products (<http://dnp.chemnetbase.com/>) and KNAPSACK (<http://kanaya.naist.jp/KNAPSACK/>) databases were used with the restraint of biological source (*Camellia sinensis*) to check if the annotated metabolites detected in our tea samples have been described before as being present in tea. In addition, we checked for phenolic compounds that were previously unambiguously identified in tea infusions using NMR<sup>20</sup> but were not present in the above mentioned *Camellia* metabolite databases.

*Data analysis:* Annotated metabolites were relatively quantified based on the peak intensity (peak height) of the specific mass signal of the parent ions. For this purpose, the raw data files were first all preprocessed (peak picking by baseline correction and noise correction) using MetAlign software<sup>21</sup> ([www.metalign.nl](http://www.metalign.nl)).<sup>22,23</sup> The 19 pre-processed data files (17 tea and 2 control samples) were used for ultrafast mass retention time searching using the Search\_LCMS module of MetAlign.<sup>24</sup> The excel compatible output contained observed accurate mass, retention time, scan number, ppm error with respect to the expected mass and intensity for each target compound in each sample. Eleven annotated compounds were missed upon this automatic peak picking, due to too close elution of isomeric species, and were therefore not taken into account. The data matrix was subsequently used for unsupervised principal components analysis (PCA) using Genemaths XT v.1.6 software ([www.applied-maths.com](http://www.applied-maths.com)). Metabolite intensities were normalized using log<sub>2</sub> transformation and standardized using range scaling. Metabolites contributing to the separation between the black and green tea samples were selected using an ANOVA ( $p < 0.01$ ). Relative quantification of 51 selected complex polyphenol conjugates (Supplemental Table 2 in the Supporting Information) was done manually using Xcalibur software and was based on the areas under the curve of the specific mass signal of the parent ions.

## Results

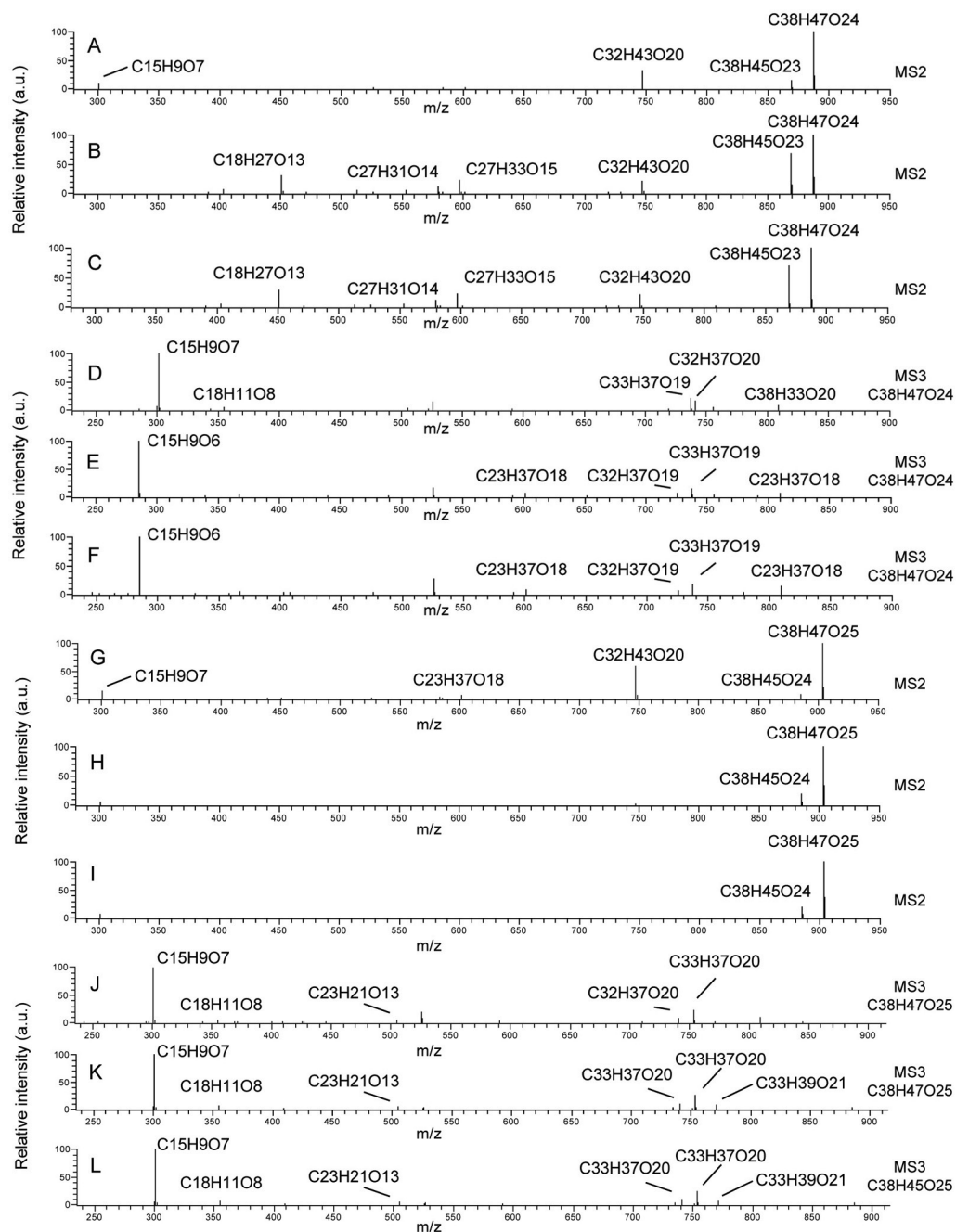
To identify phenolic compounds present in teas and profile 17 commercially available teas on the presence of the identified phenolic compounds, we extracted the tea material with aqueous methanol. Tea phenolic compounds are usually analyzed from hot water extracts of the dry leaf product. Our aim was to identify metabolites present in tea products, rather than analyzing the tea drinks. Based on preliminary results comparing aqueous-methanol extraction<sup>7,25</sup> with hot water extraction (tea drink), we decided to use the aqueous-methanol extraction, as it better represents the complete set of phenolic compounds and their conjugates present in the tea material. Moreover, the aqueous-methanol extraction protocol is a robust and general extraction method of secondary metabolites from most plant samples,<sup>22</sup> enabling the direct comparison of profiles with minimal technical variation in extracted metabolites. To test the reproducibility of this aqueous-methanol extraction for our tea samples, we performed three independent extractions for two contrasting tea products, that is, one black tea and one green tea. These extracts were subsequently analyzed by LC-Orbitrap FTMS and the peak areas and heights were determined and coefficients of variation of 10 specific tea metabolites were calculated. For all these selected metabolites, the coefficients of variation were on average well within 10% for both the black and green tea samples.

A total of 17 different aqueous methanol extracts from tea, including 7 black, 7 green, and 3 white teas, were subsequently analyzed using LC with online Orbitrap-FTMS<sup>n</sup>.<sup>18</sup> The Orbitrap FTMS was set to perform both full scan MS ( $m/z$  90-1200) and online fragmentation up to MS<sup>3</sup> for the three highest mass signals present in MS<sup>2</sup> scans. The raw data of GT China Bancha and BT Pickwick English Tea Blend teas were subsequently manually inspected, using Xcalibur software, for the presence of accurate masses (within a window of 3 ppm mass accuracy) of both parent ions and fragment ions of tea compounds previously reported in literature, which provided a good starting point for the annotation of the phenolic content of the tea metabolome. This approach resulted in the characterization of a series of 177 phenolic compounds, mainly consisting of conjugates of

the flavonols quercetin and kaempferol, and the flavan-3-ols epicatechin and catechin (Supplemental Table 1 in the Supporting Information). The MS<sup>n</sup> spectra did not reveal differences between catechin and epicatechin.<sup>26</sup> Many flavonols were substituted on their 3-hydroxyl position, which was concluded from both the presence of radical fragment ions, which are characteristic of 3-*O*-glycosides, and the fragmentation pattern of the aglycone daughter ions.<sup>18</sup> During initial screening for the quercetin aglycone fragment (C<sub>15</sub>H<sub>9</sub>O<sub>7</sub>, m/z of 301.0354 [M-H]<sup>-</sup>) in the raw LC-MS<sup>n</sup> data of the black tea extract, another compound with an almost identical mass of m/z 300.9990, corresponding to the elemental formula (EF) of C<sub>14</sub>H<sub>5</sub>O<sub>8</sub> [M-H]<sup>-</sup>, was spotted. By comparison of the fragmentation data of this molecule to available literature data we were able to identify this compound as ellagic acid.<sup>27,28</sup> In order to evaluate the relative presence of ellagic acid and the large phenolic conjugates in the different tea samples, these metabolites were manually quantified based on the area under curve of their accurate mass parent ion signal (Supplemental Table 2 in the Supporting Information).

To unambiguously elucidate the chemical structures of LC-MS<sup>n</sup> detectable metabolites, the black tea extract of Pickwick English Tea Blend and the green tea extract of China Bancha were also used for injection into an LC-TOFMS-SPE-NMR system using the compound-specific mass signal as trigger for SPE trapping of chromatographic peaks.<sup>17</sup> In this manner, we trapped 23 compounds selected from black tea and 26 compounds selected from green tea, which were subsequently eluted from the SPE cartridge using deuterated methanol and subjected to 1D-<sup>1</sup>H-NMR analysis. Different levels of metabolite identification exist and here we adopt the levels as defined by the Metabolomics Standards Initiative (MSI).<sup>29,30</sup> Tea metabolite annotations with MSI identification levels resulting from the LC-SPE-NMR approach are presented in Table 1, and additional spectral data of those metabolites can be found in Supplemental Table 1 in the Supporting Information. The combination of their MS<sup>n</sup> and 1D-<sup>1</sup>H-NMR spectral data led to the full structure identification of 33 compounds out of the 49 compounds selected for LC-SPE trapping (Table 1), including compounds previously identified in tea like epigallocatechin-3-*O*-gallate, kaempferol-3-*O*-





**Figure 1:** LC-MS<sup>n</sup> spectra of three  $C_{47}H_{54}O_{26}$  isomers (A-F) and three  $C_{47}H_{54}O_{27}$  isomers (G-L). See detailed figure caption text on next page.



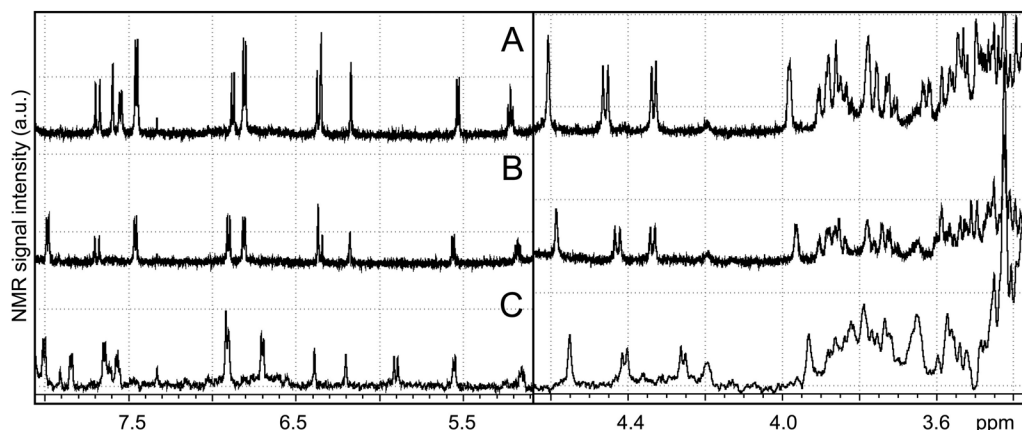
**Figure 1 (on previous page):** LC-MS<sup>n</sup> spectra of three C<sub>47</sub>H<sub>54</sub>O<sub>26</sub> isomers (A-F) and three C<sub>47</sub>H<sub>54</sub>O<sub>27</sub> isomers (G-L). The MS<sup>2</sup> and MS<sup>3</sup> spectrum of the most intense ion in MS<sup>2</sup> shown for the C<sub>47</sub>H<sub>54</sub>O<sub>26</sub> isomers at RT 38.54 (A,D), at RT 40.11 (B,E), and RT 41.12 (C,F), as well as for the C<sub>47</sub>H<sub>54</sub>O<sub>27</sub> isomers at RT 36.38 (G,J), at RT 38.11 (H,K), and RT 39.83 (I,L). The m/z values that are visible but not assigned (e.g. around m/z 525) are likely to be derived of coeluting as well as cofragmenting metabolites, as no meaningful EF could be assigned to them within 5 ppm.

**Table 1:** Compounds Identified up to MSI Level 1 or 2. Metabolites are sorted based on their m/z value followed by their retention time order. The short notation for the complex conjugates is as follows: 3-O-(2G-p-coumaroyl(trans)-3G-O-β-L-arabinosyl-3R-O-β-D-glucosylrutinoside = 3-(2G-pC-tr-3G)-2G-ara-3R-glc-rut. Table columns labels are as follows: compound names, trapped from black tea, BL (BT), and/or green tea - B (GT) extract, MSI MI level, RT in the LC-MS<sup>n</sup> system, m/z [M-H]<sup>-</sup>, EF [M-H]<sup>-</sup>, and references to literature NMR data with MeOD as solvent.

Metabolite common name	trapped from	MSI level ID	RT (min)	[M-H] <sup>-</sup> (m/z)	EF [M-H] <sup>-</sup> (CHON)	lit. data NMR
Caffeine	BT	1	17.14	193.0732	C16H13O9	
Catechin	GT,BT	1	15.61	289.0719	C15H13O6	
Epicatechin	GT,BT	1	19.53	289.0718	C15H13O6	
Gallocatechin	GT	1	9.27	305.0669	C15H13O7	
Epigallocatechin	GT	1	13.96	305.0667	C15H13O7	
p-coumaroylquinic acid-3-O trans	BT	1	15.65	337.0932	C16H17O8	
p-coumaroylquinic acid-5-O trans	BT	1	20.02	337.0932	C16H17O8	43
p-coumaroylquinic acid-4-O cis	GT	1	20.54	337.0933	C16H17O8	
p-coumaroylquinic acid-4-O trans	BT	1	21.13	337.0930	C16H17O8	
p-coumaroylquinic acid-5-O cis	BT	1	24.19	337.0930	C16H17O8	43
Theogallin (3-O-Galloylquinic) acid	BT	1	4.88	343.0673	C14H15O10	41
Epicatechin-3-O-gallate	GT	1	26.63	441.0832	C22H17O11	
Kaempferol-3-O-glucoside	BT,GT	1	31.98	447.0938	C21H19O11	17

Epigallocatechin-3-O-gallate	GT	1	21.59	457.0781	C22H17O11	
Quercetin-3-O-galactoside	BT	1	28.18	463.0886	C21H19O12	
Quercetin-3-O-glucoside	BT	1	28.67	463.0888	C21H19O12	17
Myricetin-3-O-galactoside	GT,BT	1	23.63	479.0835	C21H19O13	20
Myricetin-3-O-glucoside	BT	1	24.13	479.0836	C21H19O13	20
Kaempferol-3-O-rutinoside	BT	1	30.81	593.1518	C27H29O15	17
phloretin-3,5-di-C-glucoside	BT	1	28.97	597.1830	C27H33O15	
(Epi)gallocatechin-dimer 2	GT	2	10.67	609.1256	C30H25O14	
(Epi)gallocatechin-dimer 3	GT	2	12.77	609.1249	C30H25O14	
Quercetin-3-O-gal-1,6-rhm	BT	1	27.14	609.1464	C27H29O16	
Quercetin-3-O-glc-1,6-rhm	BT	1	27.71	609.1469	C27H29O16	17
(Epi)gallocatechin-dimer 1	GT	2	7.20	625.1249	C30H25O14	
Myricetin-3-O-(rhm-1,6-gal)	GT	1	23.53	625.1411	C27H29O17	20
Kaempferol-3-(glc-(1,6-rhm))-4'-rhm	BT	1	29.05	739.2100	C33H39O19	
Quercetin-3-O-(glc-(1,3-rhm-1,6-rhm))	GT	2	26.12	755.2049	C33H39O20	
Kaempferol-3-O-(glc-(1,3-rhm-1,6-gal))	BT	1	27.71	755.2045	C33H39O20	20
Kaempferol-3-O-(glc-(1,3-rhm-1,6-glc))	GT,BT	1	28.43	755.2044	C33H39O20	20
Quercetin-3-O-(glc-(1,3-rhm-1,6-gal))	GT,BT	1	24.84	771.1989	C33H39O21	20
Quercetin-3-O-(glc-(1,3-rhm-1,6-glc))	GT,BT	1	25.68	771.1993	C33H39O21	
Kaemp-pC-glycoside	GT	2	40.66	901.2402	C42H45O22	

Querc-3-(2G-pC-tr-3G)- 2G-ara-3R-rhm-rut	GT	1	38.54	1033.2825	C47H53O26	
Kaemp-3-(2G-pC-tr-3G)- 2G-ara-3R-glc-rut	GT	1	40.11	1033.2827	C47H53O26	
Kaemp-3-(2G-pC-cis- 3G)-2G-ara-3R-glc-rut	GT	1	41.12	1033.2826	C47H53O26	
Querc-3-(2G-pC-tr-3G)- 2G-ara-3R-glc-rut	GT	1	38.11	1049.2777	C47H53O27	32
Querc-3-(2G-pC-cis-3G)- 2G-ara-3R-glc-rut	GT	2	39.35	1049.2783	C47H53O27	
Kaemp-3-(2G-pC-tr-3G)- 2G-glc-3R-glc-rut	GT	2	38.92	1063.2933	C48 H55 O27	



**Figure 2:** Two parts of the NMR spectra (8.05–5.0 ppm, left; 4.65–3.35 ppm) of quercetin 3-*O*-(2*G*-*p*-coumaroyl(*trans*)-3*G*-*O*-*b*-*L*-arabinosyl-3*R*-*O*-*b*-*D*-glucosylrutinoside (A), kaempferol 3-*O*-(2*G*-*p*-coumaroyl(*trans*)-3*G*-*O*- $\beta$ -*L*-arabinosyl-3*R*-*O*- $\beta$ -*D*-glucosylrutinoside (B), and kaempferol 3-*O*-(2*G*-*p*-coumaroyl(*cis*)-3*G*-*O*- $\beta$ -*L*-arabinosyl-3*R*-*O*- $\beta$ -*D*-glucosylrutinoside (C). Spectrum C was processed to enhance signal intensities (line broadening = 1.5), and the signal intensities were 4 times multiplied compared to those in spectra A and B.

glucoside, and myricetin-3-*O*-galactoside, as well as *cis* and *trans* isomers of five *p*-coumaroyl quinic acid isomers which could not be differentiated based on MS<sup>n</sup> only.<sup>31</sup> Thus, NMR data was required for full identification and confirmation of the *p*-coumaroyl *trans* and *cis* isomers. In some cases, mixtures of two or more compounds were observed in the 1D-<sup>1</sup>H-NMR spectrum of a trapped LC-MS peak. However, comparison of the NMR signal integrals and using the structural

information of the LC-MS<sup>n</sup> data led to additional identification of the (minor) coeluting compounds.

We observed four isomers with an EF of C<sub>27</sub>H<sub>30</sub>O<sub>15</sub> to be present in all 17 tea samples. Two of them (eluting at retention time (RT) 29.42 and 30.85 min) could be annotated as kaempferol-3-*O*-(rhamnosyl-(1,6)-*O*-galactoside) and kaempferol-3-*O*-rutinoside (= rhamnosyl-(1,6)-*O*-glucoside), respectively, based on their MS<sup>n</sup> fragmentation patterns, elution order, and previous SPE-NMR based identification of kaempferol-3-*O*-rutinoside in tomato fruit.<sup>17</sup> Compared to these two kaempferol-*O*-glycosides, the other two isomers (RT 20.71 and 25.41 min) showed completely different fragmentation spectra. Their MS<sup>2</sup> spectra showed common fragments such as m/z of 563.1408 (EF C<sub>26</sub>H<sub>27</sub>O<sub>14</sub><sup>+</sup> [M-H]<sup>+</sup>), indicating that these two metabolites are structurally related. As their NMR spectra showed aromatic and sugar proton NMR resonances indicative of a flavonoid-C-glycoside, we could tentatively identify these compounds as apigenin-6,8-C-diglycosides.<sup>7</sup>

In the retention time window of 37-42 min, the LC-MS<sup>n</sup> data of several tea extracts revealed many doubly charged ions around m/z 510 and their corresponding singly charged ions above m/z 1000. Although the presence of these doubly charged species partly hampered fragmentation of the singly charged parent metabolites, we could obtain MS<sup>n</sup> spectra for three C<sub>47</sub>H<sub>53</sub>O<sub>26</sub> ([M-H]<sup>+</sup>) isomers (Figure 1A-1F) and three C<sub>47</sub>H<sub>53</sub>O<sub>27</sub> ([M-H]<sup>+</sup>) isomers (Figure 1G-1L). From the similarities in the MS<sup>n</sup> spectra and the small differences in elution times, we could conclude that these compounds share many structural elements. This is illustrated by spectra 1B and 1E that show similar fragmentation patterns compared to 1C and 1F, indicating that these kaempferol based conjugates are structurally highly related. From the MS<sup>n</sup> data, it can be observed that the third C<sub>47</sub>H<sub>53</sub>O<sub>26</sub> isomer (spectra 1A and 1D) is a conjugate of quercetin. The same systematic analysis was performed for the three quercetin based C<sub>47</sub>H<sub>53</sub>O<sub>27</sub> isomers (Figure 1G-1L). To fully identify these complex conjugates, we trapped them using mass-based SPE and subsequently obtained almost pure NMR spectra for 3 of them (Figure 2).

The SPE-NMR fraction of the  $C_{47}H_{53}O_{27}$  isomer eluting at RT 38.11 min yielded NMR signals that were in very good agreement with those previously reported for quercetin-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-L-arabinosyl-3*R-O-β*-D-glucosylrutinoside)<sup>32</sup> (Figure 2A). Additional two dimensional (2D)-<sup>1</sup>H-COSY NMR experiments confirmed its structure. Moreover, on the basis of the similarities observed in both MS<sup>*n*</sup> fragmentation patterns and NMR signals (Figures 1 and 2), we could unambiguously elucidate the structures of two acylated kaempferol tetraglycosides: kaempferol-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-L-arabinosyl-3*R-O-β*-D-glucosylrutinoside and its *p*-coumaroyl *cis* isomer (Figure 2B and 2C). On the basis of its RT and LC-MS<sup>*n*</sup> pattern alone, we could tentatively assign the quercetin conjugate eluting at RT 39.83 min as the *cis* isomer of quercetin-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-L-arabinosyl-3*R-O-β*-D-glucosylrutinoside.

The  $C_{47}H_{53}O_{27}$  isomer eluting at 36.38 is most likely a positional isomer of quercetin-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-L-arabinosyl-3*R-O-β*-D-glucosylrutinoside, because it yields similar fragment ions, but in different relative intensities, and upon fragmentation of the  $C_{32}H_{43}O_{20}$  fragment, the loss of its coumaroyl group was observed. On the basis of both MS<sup>*n*</sup> fragmentation patterns and diagnostic NMR signals, we identified the  $C_{47}H_{53}O_{27}$  isomer eluting at RT 38.54 min as being quercetin 3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-L-arabinosyl-3*R-O-β*-D-rhamnosylrutinoside. Furthermore, we identified kaempferol-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-β*-D-glucosyl-3*R-O-β*-D-glucosylrutinoside, because it eluted slightly earlier than its arabinose-substituted analogue and the conjugate had an observed mass of *m/z* 1063.2933 (EF  $C_{48}H_{55}O_{27}$  [M-H]). The mass difference with the arabinose-substituted analogue indeed corresponded to the mass difference of an arabinose and a glucose moiety.

The LC-MS<sup>*n*</sup> data of all acylated complex polyphenol conjugates in the tea extracts revealed two losses of  $C_9H_6O_2$  (*m/z* 146.0372) and  $C_9H_8O_3$  (*m/z* 164.0478) (Figure 1). The latter are indicative for a coumaroyl loss of complex acylated conjugates.<sup>33</sup> Analogously, the loss of  $C_9H_6O_2$  appeared to be indicative for cinnamic acid conjugation. For example, two quercetin-based acylated

glycosides were observed with  $m/z$  609.1248 (EF  $C_{30}H_{25}O_{14}$   $[M-H]^-$ ) and  $m/z$  771.1779 (EF  $C_{36}H_{35}O_{19}$   $[M-H]^-$ ), together with their kaempferol analogues. All of them were tentatively annotated based on similar fragmentation but mass difference corresponding to 1 oxygen atom and their elution order, with quercetin analogues eluting earlier than kaempferol analogues<sup>18</sup> and coumaryol analogues eluting earlier than cinnamic analogues.

Representative LC-MS chromatograms of the three different types of tea are shown in Figure 3A. The chromatogram of black tea clearly differs from those

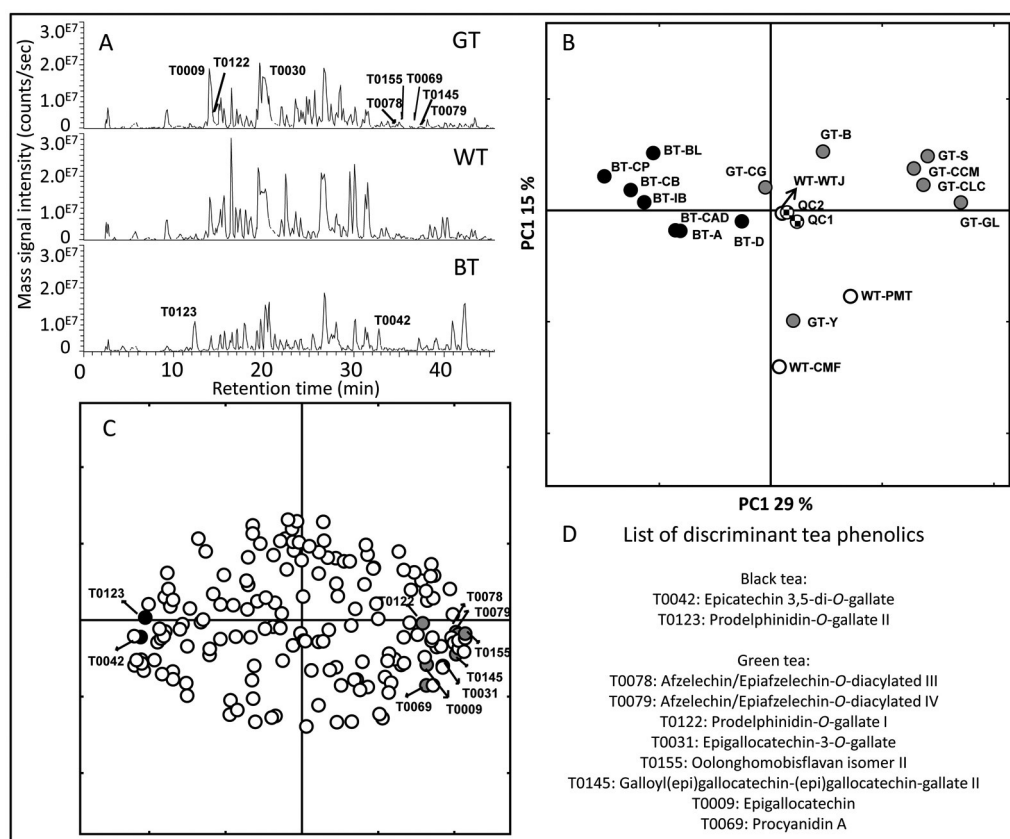


Figure 3: (A) RT 0-45 min window of three representative full-scan LC-MS profiles of green tea (GT-CG), white tea (WT-CMF), and black tea (BT-IB) with discriminative metabolites indicated. The mass signal intensities were all set to a fixed scale. Scoring (B) and loading (C) plots of unsupervised principal component analysis of 166 annotated metabolites, in which the significant discriminative metabolites are indicated in gray. (D) List of discriminative phenolic metabolites for black and green tea found in our study.

of green and white tea, and also differences exist between the white and green tea extracts. To compare the phenolic profiles and pinpoint those compounds that are mostly responsible for the differences between the 17 teas analyzed within this study, the abundance (peak height) of all annotated compounds (Table 1) was automatically extracted from the raw LC-MS data using the Search\_LCMS module of the Metalign software, on the basis of the accurate mass of the parent ions and their specific retention times. In this manner, the relative abundance of 166 out of the 177 annotated compounds could be retrieved; 11 compounds were missed as they were partly coeluting with an isomeric compound (having the same accurate mass), which did not allow automatic peak picking. The relative intensities of these 166 compounds were subsequently used as variables to perform Principal Component Analysis (PCA) on the extracts from the 17 different tea products (Figure 3B). The two technical replicates, which consisted of replicate analysis of a mix of all extracts, were positioned close to each other in the middle of the PCA plot, which indicates good technical reproducibility of the LC-MS analysis followed by the automatic peak picking and annotation procedure. The PCA plot showed a separation of, on the one hand, all 7 black teas and, on the other hand, the green and white teas (PC1, explaining 29% of the total metabolite variation). Within the group of green and white teas, a group of 4 green tea extracts clustered separately from the others, whereas the 3 white teas were spread between the green teas (Figure 3B). The metabolites that were significantly differential between the three tea types (ANOVA,  $p < 0.01$ ) are provided in Figure 3D and their presence is indicated in the chromatograms of the corresponding teas (Figure 3A). No significant differences were found between the white teas and the green teas. Eighty-two compounds were present (signal to noise ratio  $> 3$ ) in all 17 extracts and the relative abundance over the samples ranged from a factor 1.86 for gallocatechin-*O*-gallate up to a factor 2690 for quercetin-3-*O*-(rhamnosyl-(1-3)-*O*-rhamnosyl-(1,6)-*O*-glucoside, indicating large differences between teas in their levels of individual phenolic compounds.



### Discussion

Using an on-the-fly multistage accurate mass MS ( $MS^n$ ) approach in pair with LC coupled to MS-based SPE-NMR, we aimed to improve the current tea metabolome coverage by improving the identification levels of a large series of polyphenols without the need for reference compounds. We were able to annotate a total of 177 phenolic metabolites in black and green tea samples (Supplemental Table 1 in the Supporting Information). Of these annotated compounds, 82 have not been described in black and green tea extracts before, at least not at their present level of structural elucidation. We subsequently determined the variation in relative abundance of the identified phenolic compounds in 17 different tea samples, representing black, green and white teas, by using the mass specific LC-MS peak of selected large acylated conjugates (Supplemental Table 2 in the Supporting Information) as well as unsupervised multivariate analysis (PCA) based on 166 annotated phenolic metabolites. The results indicate highly variable levels in investigated tea samples for most compounds.

Full scan accurate mass is helpful in the MS analysis due to fast assignment of elemental formulas.<sup>34</sup> Moreover, mass fragmentation will help to obtain more structural information to enable annotation of metabolites to the correct compound class.<sup>18</sup> For example, Zhao et al.<sup>35</sup> observed an unknown compound of  $m/z$  499 during their UPLC-DAD-MS analysis. Our analysis using Orbitrap FTMS indicated three isomers of  $m/z$  499.0883 in the retention time region of 33 – 50 min, and their fragmentation patterns using LC- $MS^n$  led to their tentative annotation as *p*-coumaroyl-caffeoylquinic acid conjugates.<sup>18</sup> Due to relative low levels of these quinic acid conjugates in our samples and the large amounts of compounds coeluting within the specific RT region, it was not possible to obtain NMR spectra for their full identification. The power of accurate mass LC- $MS^n$  is further illustrated by the fragmentation of a GT China Chun Mee metabolite (observed although not identified by Zhao et al.<sup>35</sup>) at an  $m/z$  of 269.0455 (EF  $C_{15}H_9O_5$   $[M-H]^-$ ) that matched with the fragmentation of apigenin.<sup>17</sup>

Unequivocal identification of metabolites remains a bottleneck in many metabolomics studies,<sup>36, 37</sup> as often large numbers of tentatively identified



compounds are reported. Even though many tea metabolites have previously been fully identified, it is still hard to unambiguously annotate them. Currently, there is no single experimental database that provides all chemical knowledge of a sample, which hampers a fast identification based on a combination of observed retention time, UV absorption, MS spectra, and NMR signals. For example, the acylated quercetin tetraglycoside with an  $m/z$  of 1033.2827 ( $[M-H]^-$ ) was already identified in an oolong tea extract in 2004,<sup>32</sup> but was not fully annotated in later studies, most likely due to limited structural information.<sup>7</sup> We have previously shown that our HPLC-MS-SPE-NMR approach enabled structural elucidation of quercetin and kaempferol glycosides, including several isomers.<sup>17</sup> Indeed, in the present tea study we obtained good-quality 1D-<sup>1</sup>H-NMR data for 25 flavonoid glycosides and for three complex acylated flavonoid tetraglycosides. Nearly all flavonoids detected were specifically conjugated at the 3-O position. Only one kaempferol-triglycoside (observed  $m/z$  739.2200; EF C<sub>33</sub>H<sub>39</sub>O<sub>19</sub>  $[M-H]^-$ ) showed a different fragmentation behaviour: its triglycoside MS<sup>2</sup> spectrum displayed a fragment ion of C<sub>27</sub>H<sub>29</sub>O<sub>15</sub> that corresponds to a single deoxyhexoside substitution at either the 7-O or the 4'-O position.<sup>18</sup> Combined with its 1D-<sup>1</sup>H-NMR data, generated after LC-TOFMS-SPE trapping, its chemical structure was elucidated as kaempferol-3-*O*-(rhamnosyl-1,6-*O*-glucoside)-4'-*O*-rhamnoside. The combination of LC-MS<sup>*n*</sup> and NMR spectral data proved to be of great help in unraveling the structures of highly complex conjugated phenolics, also of compounds with  $m/z > 800$  amu (Figures 1 and 2). Although these large molecules were partly detected as doubly charged ion species by MS, our LC-MS<sup>*n*</sup> approach was able to collect sufficient MS<sup>*n*</sup> data for structural elucidation purposes. Upon combination of analytical data such as the elution order from the reversed phase column, their specific MS<sup>*n*</sup> fragmentation patterns and their characteristic proton NMR signals, a few of these large molecules could be unambiguously identified (i.e. up to MSI metabolite identification level 1). In a similar manner, we could identify (MSI level 1) 3 co-eluting compounds that showed characteristic MS<sup>*n*</sup> fragmentation patterns while only being visible as minor compounds in the 1D-<sup>1</sup>H-NMR spectra. This shows that the combination of LC-MS<sup>*n*</sup> and HPLC-

MS-SPE-NMR platforms enables the analysis of small amounts of compounds present in a highly complex extract such as tea.

For unambiguous identification we currently used 100 mg of dried tea product, which is much less than in other tea metabolite identification studies taking a few grams up to kilograms of plant sample.<sup>20, 32, 38, 39</sup> In fact, for a few metabolites, depending upon their ionization efficiency in negative electrospray ionization mode and the intensity of specific NMR signals obtained within the 15 minutes of scan time, we needed <2  $\mu$ gram of compound for full structural characterization. The success of structural elucidation of oligomers of catechin and epicatechin, as well as of theaflavins, was relatively limited in our current study. Although these compounds were clearly detectable by LC-MS and their presence in tea has been reported,<sup>4</sup> our <sup>1</sup>H-NMR measurements did not show any diagnostic aromatic signals for these phenolic molecules. Nevertheless, based on their MS<sup>n</sup> spectra, conjugated dimers of catechin or epicatechin, like Oolonghomobisflavan isomers<sup>40</sup> (observed m/z 927.1636; EF C<sub>45</sub>H<sub>35</sub>O<sub>22</sub> [M-H]<sup>-</sup>), and isomeric digalloyl substituted (epi)gallocatechin dimers (observed m/z 911.1678; EF C<sub>45</sub>H<sub>35</sub>O<sub>21</sub> [M-H]<sup>-</sup>), could be annotated. Perhaps, in order to obtain good NMR data, an SPE sorbent different from the HLB material used in the present study could be an alternative for concentration purposes, as these compounds at least showed retention on the analytical C18 HPLC column.

The LC-MS-SPE-NMR approach could trap even very polar compounds, like 3-*O*-galloylquinic acid, in a sufficient manner to obtain <sup>1</sup>H-NMR signals to compare with literature data.<sup>41</sup> Additionally, we could obtain NMR data for a series of *p*-coumaroylquinic acids. These compounds, especially the *trans* (*E*) isomers, have been previously studied by MS<sup>n</sup>, and the obtained fragmentation data matches well with those reported in literature.<sup>42-44</sup> However, the *cis* (*Z*) isomers are less well described and show the same fragmentation behavior as their corresponding (*E*) isomers.

The power of our analytical approach, using LC-MS<sup>n</sup> structural information combined with 1D-<sup>1</sup>H-NMR of MS triggered LC-SPE trapped compounds, is underlined by the fact that no time-consuming 2D-NMR spectra were necessary

to assign the diagnostic proton signals of NMR spectra of the complex acylated conjugates, even at the low total amount of complex conjugate in the NMR probe of 2 and 1  $\mu$ gram, respectively, for the quercetin and kaempferol analogues substituted by 3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O- $\beta$* -L-arabinosyl-3*R-O- $\beta$* -D-glucosylrutinoside. Moreover, the *p*-coumaroyl *cis* isomer of the kaempferol was trapped at even a much lower amount (about 0.2  $\mu$ gram); nevertheless, full identification of this molecule was possible due to the high similarity of its MS<sup>n</sup> fragmentation patterns (Figure 1) and NMR signals (Figure 2) with its trans-isomer and the observation of the diagnostic  $\alpha$  proton of the *p*-coumaroyl moiety at lower ppm resonance (5.93 vs 6.38 ppm) and a smaller coupling constant (13 Hz vs 16 Hz), as was seen in related acylated quercetin glycosides.<sup>45</sup> Literature data also reported the presence of more of these complex conjugates in tea based on LC-MS,<sup>7, 35</sup> as well as NMR.<sup>32, 38</sup> In our study, 1D-<sup>1</sup>H-NMR spectra were needed to fully unravel and confirm their structures and enable MSI level 1 annotations.

Ellagic acid was described in tea in 1941,<sup>46</sup> but has never been reported in tea samples since then until now. Ellagic acid was detected in all teas, but was present in relative higher amounts in all black teas compared to all green and white teas except two. Additionally, in all tea extracts we observed a molecule closely related to ellagic acid, which was tentatively identified as galloyl-bis-HHDP-glucose (observed *m/z* 633.0737; EF C<sub>27</sub>H<sub>21</sub>O<sub>18</sub> [M-H]). Upon MS<sup>n</sup> fragmentation of this molecule, its fragment ion C<sub>14</sub>H<sub>5</sub>O<sub>8</sub> had a similar fragmentation pattern as ellagic acid,<sup>47</sup> showing their metabolite relationship.<sup>18</sup>

Classification of different teas is greatly influenced by their phenolic content, as was reported before based on flavan-3-ols like epigallocatechin.<sup>35, 48, 49</sup> Four isomeric digalloyl substituted (epi)gallocatechin dimers and three Oolonghomobisflavan isomers were detected in all tea extracts (Supplemental Table 2 in the Supporting Information). The smallest variations in phenolic compounds between the 17 tea samples analyzed (factor 2 lower) were found for gallocatechin-3-*O*-gallate, caffeine, and epicatechin-3-*O*-gallate, suggesting that these compounds are widely distributed and do not greatly differ in their

abundance among different commercial samples. The largest variations between samples (factor 480 or higher) were observed for quercetin-3-*O*-(rhamnosyl-(1-3)-*O*-rhamnosyl-(1,6)-*O*-glucoside, kaempferol-3-*O*-(2*G-p*-coumaroyl(*trans*)-3*G-O-L*-arabinosyl-3*R-O-D*-glucosylrutinoside, and quercetin-based acylated triglycoside C<sub>42</sub>H<sub>42</sub>O<sub>22</sub> V (present in all 17 tea extracts), and for quercetin-3-*O*-glucosyl-(1-3)-*O*-rhamnosyl-(1-6)-*O*-galactoside, *p-trans*-coumaroylquinic-acid-5-*O*, and chalcane flavan-3-ol (Gambiriin) I (present in 16 out of the 17 tea extracts).

Small flavan-3-ols conjugates like epigallocatechin-3-*O*-gallate and epigallocatechin were characteristic for green tea (Figure 3). In fact, the metabolite differences that separate black and green/white teas can be correlated to changes occurring during postharvest fermentation, as is supported by previous results showing a decrease in free flavan-3-ols and their gallated conjugates during fermentation from green to black tea.<sup>50,51</sup> Interestingly, the isomers prodelphinidin-*O*-gallate II and prodelphinidin-*O*-gallate I were most characteristic for green and black tea, respectively. Unfortunately, we were not able to elucidate the structural differences between both isomeric compounds. The relative similarity between white and green teas in their metabolic composition (Figure 3B) may be partly due to the fact that both types are not fermented during after harvest.

In conclusion, our analytical approach generated highly comprehensive tea phenolic profiles and enabled annotation of 177 metabolites, of which 82 have not previously been reported at such an annotation level in black or green tea. Without the need for reference compounds and using only 100 mg of sample, we could unambiguously structurally elucidate 33 tea metabolites, using a combination of specific MS<sup>*n*</sup> fragmentation patterns and 1D-<sup>1</sup>H-NMR signals. In the present study, the identified metabolites were used to compare the phenolic profiles of 17 tea samples and to pinpoint differential metabolites. We are confident that these advanced metabolite identification approaches will increase our knowledge of the differences in chemical composition tea samples, for instance, in relation to genotype, growth conditions, postharvest processing, and preparation of tea drinks, as well as chemical profiles of material from other biological sources.

## *Safety*

There were no special safety concerns for this study and the authors declare that all chemicals in this study were used according their safety precautions.

## *Supporting Information description*

*Supporting Table 1.* List of 177 annotated metabolites. In case that NMR spectra were obtained (column D, SPE-NMR), additional spectral data is provided (columns L-O). Metabolites were sorted based on their accurate mass ( $[M-H]^-$  ( $m/z$ ), column G), and finally on their retention times (RT Orbitrap, column E).

*Supporting Table 2.* List of 51 metabolites that were relatively quantified in all 17 tea extracts based on their area under the curve of the specific parent mass signal and their relative areas in the different samples. Metabolites were sorted based on their accurate mass ( $[M-H]^-$  ( $m/z$ ), column D), and finally on their retention times (RT Orbitrap, column B).

This material is available free of charge via the Internet at <http://pubs.acs.org>. The supplemental Excel file is available at <http://edepot.wur.nl/216854> and a direct links are: (Supporting Table 1) <http://edepot.wur.nl/216859> and (Supporting Table 2) <http://edepot.wur.nl/216860>

## *Funding sources*

This research was granted by the Netherlands Metabolomics Centre and the Centre for BioSystems Genomics, both of which are part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

## *Notes*

The authors declare no competing financial interest.

## *Abbreviations Used*

BT, black tea; DAD, diode array detector; EF, Elemental Formula; FT, Fourier transformed; GAL, galactose moiety; GLC, glucose moiety; GT, green tea; HHDP, 6,6'-dicarbonyl-2,2',3,3',4,4'-hexahydroxybiphenyl moiety;  $m/z$ , mass to charge; MS, Mass Spectrometry; NMR, Nuclear Magnetic Resonance; MI, Metabolite Identification; MSI, Metabolomics Standards Initiative; PCA, Principal Component Analysis; TOF, time of flight; RT, retention time; RHM, rhamnose moiety; UV, ultraviolet; WT, white tea.

## Acknowledgements

The authors would like to thank the tea shop Delicious in Hoogland, The Netherlands, for providing batches of tea. We also thank Dr. Saiko Yoshida for her help.

## References

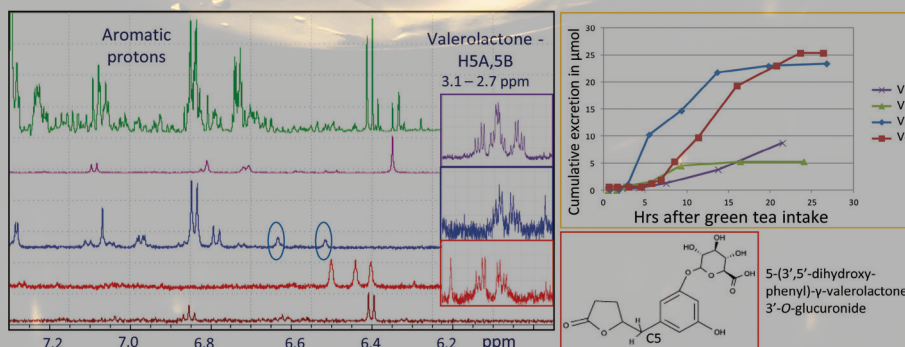
1. S. Sang, J. D. Lambert, C.-T. Ho and C. S. Yang, *Pharmacol. Res.*, 2011, **64**, 87-99.
2. J. M. Hodgson and K. D. Croft, *Mol. Asp. of Med.*, 2010, **31**, 495-502.
3. C. S. Yang, H. Wang, G. X. Li, Z. Yang, F. Guan and H. Jin, *Pharmacol. Res.*, 2011, **64**, 113-122.
4. J. Peterson, J. Dwyer, S. Bhagwat, D. Haytowitz, J. Holden, A. L. Eldridge, G. Beecher and J. Aladesanmi, *J. Food Comp. Anal.*, 2005, **18**, 487-501.
5. N. Kuhnert, J. W. Drynan, J. Obuchowicz, M. N. Clifford and M. Witt, *Rapid Commun. Mass Spectrom.*, 2011, **24**, 3387-3404.
6. J. W. Drynan, M. N. Clifford, J. Obuchowicz and N. Kuhnert, *Nat. Prod. Rep.*, 2010, **27**, 417-462.
7. L.-Z. Lin, P. Chen and J. M. Harnly, *J. Agr. Food Chem.*, 2008, **56**, 8130-8140.
8. C. Manach, J. Hubert, R. Llorach and A. Scalbert, *Mol. Nutr. Food Res.*, 2009, **53**, 1303-1315.
9. M. Bedair and L. W. Sumner, *Tr. Anal. Chem.*, 2008, **27**, 238-250.
10. D. Del Rio, A. J. Stewart, W. Mullen, J. Burns, M. E. J. Lean, F. Brighenti and A. Crozier, *J. Agr. Food Chem.*, 2004, **52**, 2807-2815.
11. D. Wang, J. Lu, A. Miao, Z. Xie and D. Yang, *J. Food Comp. Anal.*, 2008, **21**, 361-369.
12. C. Wu, H. Xu, J. Héritier and W. Andlauer, *Food Chem.*, 2012, **132**, 144-149.
13. J. Dou, V. S. Y. Lee, J. T. C. Tzen and M.-R. Lee, *J. Agr. Food Chem.*, 2007, **55**, 7462-7468.
14. J. Peterson, J. Dwyer, P. Jacques, W. Rand, R. Prior and K. Chui, *J. Food Comp. Anal.*, 2004, **17**, 397-405.
15. Y. Wang, Q. Li, Q. Wang, Y. Li, J. Ling, L. Liu, X. Chen and K. Bi, *J. Agr. Food Chem.*, 2011.
16. H. Karaköse, R. Jaiswal and N. Kuhnert, *J. Agr. Food Chem.*, 2011, **59**, 10143-10150.
17. J. J. J. Van der Hoof, V. Mihaleva, R. J. Bino, R. C. H. de Vos and J. Vervoort, *Magn. Res. Chem.*, 2011, **49**, S55-S60.
18. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino and R. C. H. de Vos, *Metabolomics*, 2012, **8**, 691-703.
19. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and R. C. H. de Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.
20. S. Scharbert, N. Holzmann and T. Hofmann, *J. Agr. Food Chem.*, 2004, **52**, 3498-3508.
21. A. Lommen, *Anal. Chem.*, 2009, **81**, 3079-3086.
22. R. C. H. de Vos, S. Moco, A. Lommen, J. J. Keurentjes, R. J. Bino and R. D. Hall, *Nat. Prot.*, 2007, **2**, 778-791.
23. Y. M. Tikunov, R. C. H. de Vos, A. M. G. Paramás, R. D. Hall and A. G. Bovy, *Plant Physiol.*, 2010, **152**, 55-70.
24. A. Lommen, A. Gerssen, J. E. Oosterink, H. J. Kools, A. Ruiz-Aracama, R. J. B. Peters and H. G. J. Mol, *Metabolomics*, 2011, **7**, 15-24.
25. G. Le Gall, I. J. Colquhoun and M. Defernez, *J. Agr. Food Chem.*, 2004, **52**, 692-700.

26. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino, J. Beekwilder and R. C. H. De Vos, *Anal. Chem.*, 2011, **83**, 409-416.
27. J.-H. Lee, J. V. Johnson and S. T. Talcott, *J. Agr. Food Chem.*, 2005, **53**, 6003-6010.
28. W. Mullen, T. Yokota, M. E. J. Lean and A. Crozier, *Phytochemistry*, 2003, **64**, 617-624.
29. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211-221.
30. R. J. Molyneux and P. Schieberle, *J. Agr. Food Chem.*, 2007, **55**, 4625-4629.
31. M. N. Clifford, J. Kirkpatrick, N. Kuhnert, H. Roozendaal and P. R. Salgado, *Food Chem.*, 2008, **106**, 379-385.
32. R. Mihara, T. Mitsunaga, Y. Fukui, M. Nakai, N. Yamaji and H. Shibata, *Tetrahedron Lett.*, 2004, **45**, 5077-5080.
33. B. Harbaum, E. M. Hubbermann, C. Wolff, R. Herges, Z. Zhu and K. Schwarz, *J. Agr. Food Chem.*, 2007, **55**, 8251-8260.
34. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23-60.
35. Y. Zhao, P. Chen, L. Lin, J. M. Harnly, L. Yu and Z. Li, *Food Chem.*, 2011, **126**, 1269-1277.
36. S. Moco, J. Vervoort, R. J. Bino, R. C. H. De Vos and R. Bino, *Tr. Anal. Chem.*, 2007, **26**, 855-866.
37. A. Scalbert, C. Andres-Lacueva, M. Arita, P. Kroon, C. Manach, M. Urpi-Sarda and D. Wishart, *J. Agr. Food Chem.*, 2011, **59**, 4331-4348.
38. V. S.-Y. Lee, C.-R. Chen, Y.-W. Liao, J. T.-C. Tzen and C.-I. Chang, *Chem. Pharmaceut. Bull.*, 2008, **56**, 851-853.
39. J.-h. Lu, Y. Liu, Y.-y. Zhao and G.-z. Tu, *Chem. Pharmaceut. Bull.*, 2004, **52**, 276-278.
40. F. Hashimoto, G. I. Nonaka and I. Nishioka, *Chem. Pharmaceut. Bull.*, 1989, **37**, 3255-3263.
41. H. Nishimura, G.-I. Nonaka and I. Nishioka, *Phytochemistry*, 1984, **23**, 2621-2623.
42. M. N. Clifford, K. L. Johnston, S. Knight and N. Kuhnert, *J. Agr. Food Chem.*, 2003, **51**, 2900-2911.
43. Y. Lu, Y. Sun, L. Y. Foo, W. C. McNabb and A. L. Molan, *Phytochemistry*, 2000, **55**, 67-75.
44. A. Plazonić, F. Bucar, Ž. Maleš, A. Mornar, B. Nigović and N. Kujundžić, *Molecules*, 2009, **14**, 2466-2490.
45. J. G. Diaz and W. Herz, *Phytochemistry*, 2010, **71**, 463-468.
46. M. Tsujimura, *Sc. Pap. I. P. C. R.*, 1941, **38**, 487-489.
47. M. Kajdžanoska, V. Gjamovski and M. Stefova, *Macedon. J. Chem. Chem. Engineer.*, 2010, **29**, 181-194.
48. J.-H. Li, A. Nesumi, K. Shimizu, Y. Sakata, M.-Z. Liang, Q.-Y. He, H.-J. Zhou and F. Hashimoto, *Phytochemistry*, 2011, **71**, 1342-1349.
49. G. Xie, M. Ye, Y. Wang, Y. Ni, M. Su, H. Huang, M. Qiu, A. Zhao, X. Zheng, T. Chen and W. Jia, *J. Agr. Food Chem.*, 2009, **57**, 3046-3054.
50. J.-E. Lee, B.-J. Lee, J.-O. Chung, H.-J. Shin, S.-J. Lee, C.-H. Lee and Y.-S. Hong, *Food Res. Int.*, 2011, **44**, 597-604.
51. K. M. Ku, J. Kim, H.-J. Park, K.-H. Liu and C. H. Lee, *J. Agr. Food Chem.*, 2010, **58**, 345-352.





## Chapter 6: Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake



Justin J.J. van der Hooft, Ric C.H. de Vos, Raoul J. Bino,  
Velitchka Mihaleva, Lars Ridder, Niels de Roo, Doris M. Jacobs,  
John P.M. van Duynhoven, and Jacques Vervoort

This chapter was published as research article in *Analytical Chemistry*, 2012, volume 84 (16), pp 7263-7271, DOI: 10.1021/ac3017339

### *Abstract*

In dietary polyphenol exposure studies, annotation and identification of urinary metabolites present at micromolar concentrations remain major obstacles. In order to identify potential bioactive components, it is necessary to have the exact structural elucidation and quantification of polyphenol-derived conjugates that are present in the human body. In order to identify and quantify metabolites and conjugates excreted in human urine after single bolus intake of black or green tea, the present study applied a combination of a solid phase extraction (SPE) preparation step and two high pressure liquid chromatography (HPLC)-based analytical platforms; namely, accurate mass fragmentation (HPLC-FTMS<sup>n</sup>) and mass-guided SPE-trapping of selected compounds for NMR measurements (HPLC-TOFMS-SPE-NMR). HPLC-FTMS<sup>n</sup> analysis led to the annotation of 138 urinary metabolites, including 48 valerolactone and valeric acid conjugates. By combining the results from MS<sup>n</sup> fragmentation with the one dimensional (1D)-<sup>1</sup>H-NMR spectra of HPLC-TOFMS-SPE trapped compounds, we elucidated the structures of 36 phenolic conjugates, including the glucuronides of 3',4'-di, and 3',4',5'-trihydroxyphenyl- $\gamma$ -valerolactone, three urolithin glucuronides, and indole-3-acetic acid glucuronide. We also obtained 26 hours of quantitative excretion profiles for certain valerolactone conjugates. The study concludes that the combination of HPLC-FTMS<sup>n</sup> and HPLC-TOFMS-SPE-NMR platforms results in the efficient identification and quantification of low abundant phenolic conjugates down to nanomoles of trapped amounts of metabolite corresponding to micromolar metabolite concentrations in urine.

## *Introduction*

Polyphenols are a class of functional ingredients that have been widely studied, due to their potential health benefits.<sup>1-3</sup> There is increasing awareness that these beneficial effects could be partly caused by breakdown products of polyphenols formed in the gut.<sup>4-6</sup> In order to gain greater insight into the metabolic fate of polyphenols, it is necessary to identify the core structures and conjugation patterns of gut microbial metabolites.<sup>1, 7, 8</sup> Tea is a rich source of flavonoids, which are mainly represented in green tea by (epi)catechin, (epi)gallocatechin (flavan-3-ols), and galloylated conjugates thereof, as well as oxidized products in black tea.<sup>8</sup> Following tea intake, food components like flavan-3-ols and their conjugates arrive in the gastro-intestinal tract. A small part of these components is absorbed and modified in the small intestine, while the major part reaches the large intestine and is subject to colonic microbial breakdown, followed by absorption in the bloodstream or excretion in the faeces.<sup>9-11</sup> Once absorbed, the microbial breakdown products reach the liver. Here, phase II conjugation reactions can occur, which lead to glucuronidated, sulphated, and methylated conjugates or a combination thereof.<sup>2</sup> These conjugates then circulate in the blood, may undergo enterohepatic cycling, and then finally leave the body by excretion in the urine.<sup>2</sup> Five factors hamper the identification of conjugates in human urine: (i) the complex composition of urine, including different metabolite classes at a wide range of concentrations, for example the high-abundant (millimolar range) hippuric acid;<sup>12</sup> (ii) the complex matrix of urine, caused by the presence of high concentrations of salt and urea; and (iii) the low concentration of the conjugates, i.e., micromolar concentrations in urine pooled for 24 hours;<sup>13-15</sup> and (iv) the many possible isomeric forms of the metabolites that are present; and finally (v) the absence of reference compounds that allow for MS based identification and quantification.<sup>16, 17</sup> As a result of these obstacles, very few NMR spectra of intact conjugated breakdown products are currently available as reference data, as most dietary polyphenol exposure studies enzymatically deconjugated the sulphated and glucuronidated polyphenol core or its phenolic microbial breakdown products before data acquisition and analysis. To partly overcome

these obstacles, the present study used a solid phase extraction (SPE) sample preparation procedure that both purified and concentrated the (poly)phenolic metabolites.<sup>18, 19</sup>

So far, mainly high pressure liquid chromatography (HPLC) - Diode Array Detection (DAD) and HPLC - Mass Spectrometry (MS) approaches have been used to characterize breakdown products of dietary polyphenols in urine.<sup>8</sup> In most of these studies the circulating species are first enzymatically deconjugated before analysis, thereby increasing the concentrations of the core structures and at the same time reducing the complexity of the sample. Studies on intact conjugates reported the presence of several series of isomeric conjugates, due to the differences in phenolic core structures and the presence of multiple conjugation sites.<sup>20-22</sup> While MS/MS and MS<sup>n</sup> fragmentation can be very helpful in metabolite characterization and identification,<sup>23-25</sup> it is usually impossible to determine a substitution position solely based on MS data, unless reference compounds are available.<sup>26</sup> The additional structural information from nuclear magnetic resonance spectroscopy (NMR) is then needed to elucidate the complete structure.<sup>27</sup>

The metabolite content of tea has been well-studied over the last years.<sup>8, 28, 29</sup> Black tea mainly contains theaflavins and thearubigins, oxidized products of the flavan-3-ols epicatechin and catechin and their galloylated conjugates that are abundantly present in green tea. A large range of quercetin and kaempferol glycosides can be found in tea, mostly glycosylated on the 3-*O* position.<sup>30, 31</sup> The fate of these tea polyphenols in the human body as a result of uptake in the small and large intestine has been subject of study over the latest years.<sup>8, 32, 33</sup> Several gut microbial derived compounds such as hippuric acid and pyrogallol-2-*O*-sulphate have been found in elevated amounts in urine based on acquired 1D-<sup>1</sup>H-NMR spectra after black tea intake.<sup>2, 12, 34</sup>

Direct 1D-<sup>1</sup>H-NMR analysis of crude urine extracts can result in spectral data that are difficult to interpret, due to the presence of overlapping NMR signals in the spectra of its complex metabolome as well as dominating NMR signals of highly abundant metabolites, such as hippuric acid.<sup>12, 35</sup> These dominant

signals hinder the identification of the lower abundant intact phenol conjugates. Therefore, we used the recently developed MS based SPE trapping procedure which was coupled to  $^1\text{H-NMR}$  structural identification.<sup>27</sup> This approach led to (partially) purified compounds with 1D- $^1\text{H-NMR}$  spectra that were used for metabolite identification of plant derived metabolites.<sup>27, 31</sup>

Here, we report a methodology that combines i) a SPE-based sample preparation step for concentration and purification of the urinary metabolites, and ii) HPLC-FTMS of the intact metabolites followed by HPLC-TOFMS guided SPE-NMR of selected compounds for their structural elucidation. Using this approach, we achieved the identification and quantification of intact metabolite conjugates present in human urine after a single bolus intake of black or green tea.

### *Material and Methods*

*Chemicals:* Acetonitrile (HPLC grade) was obtained from Biosolve (Valkenswaard, The Netherlands), acetonitrile NMR Chromasolve (for HPLC-NMR; SPE device) from Riedel-de Haën (Seelze, Germany), methanol (HPLC grade) from Merck-Schuchardt (Hohenbrunn, Germany), phosphoric acid (85–88 % pure) from Sigma-Aldrich, (Steinheim, Swiss), and formic acid (99–100 % pure) from VWR international S.A.S. (Briare, France). Deuterated methanol (99.8% pure) was obtained from CIL Inc., (Andover, MA). Ultrapure water was made in purification units present in-house.

*Tea intake protocol:* We used a protocol for black tea and for green tea intake based on previous phenolic intake studies in literature.<sup>35</sup> In the black tea study, two healthy male volunteers (28 and 55 years) consumed, after one and a half day of polyphenol-low diet (i.e., consuming no coffee, tea, wine, fruit juices, and cereal bread at all and as few as possible vegetables and fruit), 1.5 L of black tea without sugar, milk or other additives (6 grams of Pickwick English Blend tea extracted with 1.5 L boiling water for 5 minutes) within two hours in the morning. During 24 hours, 50 ml of urine fractions were collected upon spontaneous urination for the initial testing of the analytic approach. Immediately after collection, urine samples were stored at 4 °C for maximal 24 hours before

transferring to -80 °C.

In the green tea study, four healthy male volunteers (24 – 55 years) followed the polyphenol-low diet for 3 days, to ensure wash-out of previously dietary phenol metabolites, prior to green tea intake. Volunteers consumed 1.5 L of green tea (6 grams of China Bancha Green tea extracted with 1.0 L hot water for 5 minutes). Urine samples were collected in 50 ml tubes until 26 hours after intake. Immediately after collection, the volume of the sample was recorded, and the urine samples were stored at 4 °C for maximal 24 hours before transferring to -80 °C. The protocols were approved by the Medical Ethics Review Committee (METC) of Wageningen University.

*Sample preparation steps.*

*Solid Phase Extraction protocol:* HLB solid phase extraction (SPE) cartridges (3cc, OASIS, Waters) were washed and activated with 4 ml of methanol and conditioned with 6 ml of H<sub>2</sub>O under vacuum. Then, 3 ml of urine sample was diluted 1:1 with H<sub>2</sub>O containing 4% phosphoric acid (final pH about 3.0). The 6 ml was put on the SPE cartridges, followed by a washing with 4 ml of H<sub>2</sub>O. Elution was performed with 4 ml 100% methanol. Subsequently, the eluate was freeze-dried over night in a speedvac protected from light at 35 °C and stored at -20 °C before further handling.

*Further sample preparation:* The freeze-dried samples were redissolved in 200 µL of appropriate solvent for MS or NMR measurements. For SPE trapping, a 25% MeOH in water containing 0.1% formic acid was used to minimize chromatographic drifting of the relative polar conjugates. In detail sample preparation for *1D-<sup>1</sup>H-NMR*, *HPLC-FTMS<sup>n</sup>*, *HPLC-TOFMS-SPE-NMR*, and *Offline-FTMS<sup>n</sup>* measurements can be found in Supplemental Text 1.

*Analytical details.*

*HPLC-PDA-FTMS<sup>n</sup> conditions:* HPLC-FTMS<sup>n</sup> protocols were used as described earlier<sup>25</sup> with some adaptations that are listed in Supplemental text 1.

*1D-<sup>1</sup>H-NMR experiments of the SPE enriched urine fractions:* The 200 µL

SPE enriched urine fractions in MeOD were transferred into 3 mm NMR tubes and placed in a 600 MHz Bruker equipped with a cryoprobe, using a Bruker sample handler (BACS-60) and run in automatic mode, with the following main parameters for NOESY 1D-<sup>1</sup>H-NMR acquisition: number of scans of 256, relaxation time (d1) of 4 s, receiver gain of 128, spectral width of 30 ppm, time domain of 97352 points, using water suppression, at a temperature of 300 K. NMR spectra were base-line corrected, phase corrected, and calibrated on the residual MeOH-d<sub>3</sub> solvent peak.

*HPLC-TOFMS-SPE-NMR conditions:* The HPLC-TOFMS-SPE-NMR (Bruker) set up was used as described in van der Hooft et al. (2011).<sup>27</sup> In short, an Agilent 1200 series HPLC tower is coupled to a MicroTOF MS and a Prospekt unit that contains 2.1 mm SPE cartridges. To ensure stable chromatography of the relative polar conjugates a pre-run equilibration time of 20 minutes was applied, followed by a gradient of 0.1% formic acid in water (solvent A) and acetonitrile containing 0.1% formic acid (solvent B): at 0 minutes 3% B, at 5 minutes 3% B, at 25 minutes 7% B, at 45 minutes 20% B, at 60 minutes 50% B, at 70 minutes 50% B, at 73 minutes 3% B, and a 10 minutes conditioning step of 3% B. For black tea-derived urine compounds, 5 cumulative trappings of 28 metabolites, including base peaks and peaks selected based on their mass spectral features were performed with an injection volume of 30 μL, while for the green tea-derived urine compounds 5 cumulative trappings of 40 μL injection volume were performed for 25 selected LC-MS peaks. The cartridges with the trapped LC-MS peaks were dried with nitrogen gas and the analytes were eluted with deuterated methanol (MeOD) into the NMR flow probe. <sup>1</sup>H NMR spectra were recorded as follows: typically a test run of 8 scans was performed followed by a 128 scans experiment (about 15 minutes measurement time) with the same basic parameters as mentioned for the 1D-<sup>1</sup>H-NMR experiments of the SPE enriched urine samples. If needed for further structural confirmation and given sufficient analyte was present (signal to noise ratio was more than 8), 2D-NMR measurements (COSY, TOCSY, HSQC, and HMBC experiments) were performed as well. All SPE trapped analytes were collected in Eppendorf



tubes and stored at -20 °C.

*Separation of urolithin conjugates:* In order to separate the observed glucuronides of urolithin A and B, an adapted HPLC gradient was applied: 5% B from 0 to 20 minutes, then a linear increase to 15% B in 25 minutes, and to 50% at another 25 minutes, followed by washing and equilibration. Using this gradient, 13 compounds, including urolithin A and B glucuronides, from urine were selected for 4 cumulative trappings with 40 µl injection volume.

*Offline FTMS<sup>n</sup> of collected SPE trapped analytes:* In order to obtain in-depth spectral trees, offline MS<sup>n</sup> in both negative and positive ionization mode was applied to selected m/z values of trapped LC-MS peaks using protocols as earlier described with an adapted selection width of 1.5 Da for the precursor ion.<sup>36</sup>

#### *Data analysis.*

*Structure confirmation with the PERCH NMR software:* The structure of the conjugates was confirmed by comparison of the experimental chemical shifts and couplings with those predicated with the PERCH NMR software.<sup>27,37,38</sup> The three dimensional (3D) structures of all deconjugated and all glucuronidated and sulphated isomers were generated for each phenylvalerolactone (using the 4R stereoconfiguration).<sup>39</sup> The selection of the correct isomer was based on the aromatic protons of the phenyl moiety. In the conjugates, the largest change in the chemical shifts of proton NMR signals when compared to signals of the deconjugated compound is observed for the protons adjacent to the glucuronate and sulphate groups. The selected isomer had the best match between experimental and predicted chemical shifts. When pure 1D-<sup>1</sup>H-NMR spectra were available, these were fitted using the PERCH NMR software. The signal of formic acid (at 8.1 ppm in MeOD) was also included in the fit procedure so that the signals in the 8.3-8.0 ppm range were fully resolved.

*Metabolite identification:* Metabolites were annotated as novel if no hit of the completely elucidated structure was obtained from HMDB (<http://www.hmdb.ca/>), Scifinder (<https://scifinder.cas.org/scifinder/>) or recently reported studies on urinary phenolic conjugates.<sup>21,22</sup> Queries were based on assigned elemental

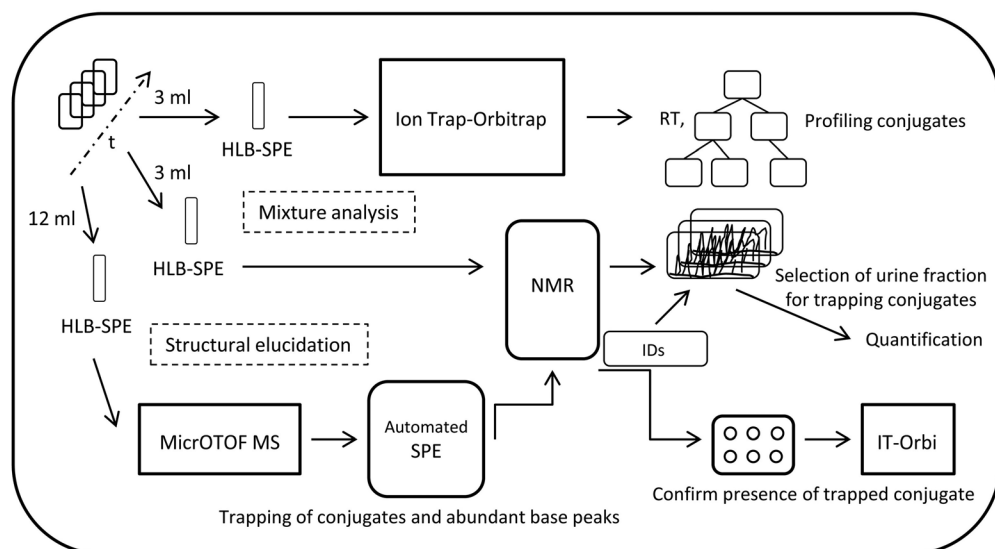


formulas (EFs) using a 3 ppm threshold in Xcalibur. If known and needed to reduce the number of retrieved candidate structures, the core structure of the conjugate was used as restraint.

*Metabolite quantification:* In order to quantify metabolites in urine, we externally calibrated the NMR using *in triplo* measurements of three concentrations (5, 50, and 400  $\mu\text{gram/ml}$ ) of rutin (quercetin-3-*O*-rutinoside), in 3 mm NMR tubes. Calibration lines were drawn based on the integrals of the A-ring proton (6.07 ppm, H6), B-ring proton (7.65 ppm, H2' and H6'), and the anomeric proton of rhamnose (4.52 ppm, H1''') with proton assignments as in van der Hooft et al., 2011.<sup>27</sup> Due to small differences in relaxation times for these protons and the possibility of intensity reduction of the anomeric proton due to the suppression of the nearby water signal, it was decided to use the B-ring proton peak area for quantification. Before acquisition, automatic tuning and matching, shimming, and 90° pulse calculation (TopSpin) was performed, which resulted in a small range of 90° pulse ( $9.2 \pm 0.15$  ms) for both 3 mm tubes and the flow probe insert. For quantification calculations, an average SPE recovery of 85% was used (see Results section, supplemental figure 1). The determined concentrations in  $\mu\text{M}$  in each urine fraction were converted into excreted  $\mu\text{moles}$  after 1.5 L tea intake using the recorded urine fraction volumes. It should be noted here that overlap with <sup>1</sup>H-NMR signals from other metabolites is a possible source of error (depending on local crowdedness of spectra from a few % up to 20%) for this approach. The obtained quantification values should therefore be regarded as a maximum concentration.

## *Results*

A combined approach of SPE sample preparation, HPLC-FTMS<sup>n</sup> and HPLC-TOFMS-SPE-NMR was applied to urine samples individually collected within 26 hours after consumption of black or green tea (Figure 1). Two urine fractions sampled at 12 hours after black tea intake were used for a proof of concept experiment to explore the experimental setup for metabolite annotation and identification, followed by the green tea intake experiment in which also



**Figure 1:** Extended analytical approach in schematic view – Urine fractions were collected at different time points after green tea consumption and were subjected to an SPE procedure before analysis by i) HPLC-FTMS<sup>n</sup> and ii) 1D-<sup>1</sup>H-NMR. From these data sets, urine fractions containing highest levels of target compounds were selected for HPLC-TOFMS-SPE-NMR in order to structurally elucidate their conjugates. The obtained SPE fractions were injected into the FTMS using direct-infusion MS<sup>n</sup> for confirmation of trapped masses and in-depth MS<sup>n</sup> spectral tree generation. Diagnostic <sup>1</sup>H-NMR signals were used to quantify identified conjugates in the urine fractions.

quantitative profiling of metabolite conjugates in four volunteers over the 26 hours sampling time was included.

**SPE sample preparation:** We used an improved version of a previously described HLB-SPE sample preparation procedure to concentrate the urine conjugates 15 fold and to remove part of the complex background.<sup>19</sup> Initially, a recovery test of the SPE extraction procedure was performed *in triplo* using a urine extract containing black tea-derived conjugates and analyzing the recovery of 17 conjugates by HPLC-TOFMS. This resulted in an average recovery of 85% for 15 out of 17 conjugates based on peak heights in HPLC-TOFMS (Suppl. Fig. S1). From this result, we concluded that our SPE procedure was efficient in retaining and eluting urine conjugates using methanol as eluents.

**1D-<sup>1</sup>H-NMR measurements of urine extracts:** To analyze the presence of NMR

detectable phenolic metabolites in urine after green tea consumption, 1D-<sup>1</sup>H-NMR spectra of all urine SPE concentrated extracts were obtained. During initial analysis, the NMR spectra were used to screen for urine fractions containing a significant amount of signals possibly belonging to tea-derived conjugates, for example anomeric protons of a glucuronide (i.e. glucuronic acid) moiety, typically appearing in the 4.8 – 6.0 ppm region as a doublet of 7.8 Hz, as well as NMR signals in the aromatic region (6.0 – 7.0 ppm). We then selected three urine fractions (collected between 5 and 11 hours tea after intake) for detailed HPLC-FTMS<sup>n</sup> analysis and for SPE trapping.

*HPLC-FTMS<sup>n</sup> analysis of urine extracts.*

*HPLC-FTMS<sup>n</sup> metabolite annotation:* To annotate possible metabolites and metabolite breakdown products in urine after green tea intake, we used HPLC-FTMS<sup>n</sup>.<sup>25</sup> From literature, a shortlist was created of possible tea breakdown products and their conjugates present in human urine.<sup>21, 22, 40</sup> The reported masses and their fragmentations were compared to the HPLC-FTMS<sup>n</sup> spectra actually obtained from the green tea urine extracts. Glucuronide conjugates were recognized by the neutral loss of 176.0321 Da (corresponding to an elemental formula (EF) of C<sub>6</sub>H<sub>8</sub>O<sub>6</sub>) and the glucuronide moiety ion itself, i.e., m/z 175.0248 with elemental EF C<sub>6</sub>H<sub>7</sub>O<sub>6</sub> [M-H]<sup>-</sup>, and its fragments, i.e., m/z 113.0244 with EF C<sub>5</sub>H<sub>5</sub>O<sub>3</sub> [M-H]<sup>-</sup>, and m/z 151.0142 with EF C<sub>6</sub>H<sub>5</sub>O<sub>5</sub> [M-H]<sup>-</sup>. Sulphate conjugates were characterized by the neutral loss of 79.9568 Da (corresponding to EF SO<sub>3</sub>), and the presence of SO<sub>3</sub><sup>-</sup> in the fragmentation spectra (i.e. m/z 79.9573). Methylated metabolites could often be recognized by a loss of 15.0234 Da (CH<sub>3</sub>). In urine samples that were derived after black tea intake, the accurate mass of deprotonated vanillic acid (i.e. m/z 167.0350) yielded 2 glucuronide conjugates and one sulphate conjugate in the HPLC-FTMS<sup>n</sup> chromatogram. The HPLC-FTMS<sup>n</sup> data also indicated the presence of the glycine conjugates hippuric acid, vanilloyl glycine, and phenylacetyl glycine (PAG) in substantial amounts. In the urine of 7 hours after green tea intake, accurate masses of valerolactone and valeric acid conjugates were detected,<sup>22</sup> of which the identity was confirmed

by MS/MS characteristics reported earlier.<sup>21, 22</sup> This led to the annotation of 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone, 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone, and 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone core structures. In addition, 9 valerolactone conjugates, including O-methylated ones, were annotated for the first time. Following this approach, HPLC-FTMS<sup>n</sup> profiling of the SPE-concentrated extracts of urine after black and green tea consumption resulted in the annotation of 138 urinary metabolites, of which 25 were valerolactone conjugates and 23 were valeric acid conjugates (Supplemental Table 1). Their annotations reached (solely based on HPLC-FTMS<sup>n</sup> data) mainly Metabolite Standards Initiative (MSI) metabolite identification (MI) level 2 and 3.<sup>41</sup>

*HPLC-FTMS based relative quantification of conjugates:* The cumulative urinary excretion of annotated metabolites over the 26 hours sampling was derived from the peak areas in full scan filtered HPLC-FTMS<sup>n</sup> profiles (Suppl. Fig S2). Differences in nutrkinetic parameters in-between metabolites and in-between volunteers can be derived from the plots, such as differential lag times, high and low responders, as well as slow and fast responders.<sup>35, 42</sup> For example, methylgallic acid and (epi)catechin-*O*-sulphate-*O*-methyl conjugates reached their total cumulative excretion 4-5 hours after green tea intake (Suppl. Fig. S2A). The cumulative excretion of valerolactone and valeric acid conjugates, however, leveled off at 8-11 hours after green tea intake and their excretion was detected in urine fractions up to 26 hours after intake (Suppl. Fig. S2A). These results are in line with earlier observations for (epi)catechin metabolites and microbial derived metabolites after a polyphenol rich beverage intake.<sup>43</sup> In Figure S2B-S2G, the HPLC-FTMS based cumulative excretion of 5 valerolactone conjugates and 1 valeric acid conjugate in the urine fractions of the 4 volunteers is plotted. The plots show similar excretion profiles of conjugates within one volunteer, but also differences between the volunteers for these conjugated microbial metabolites. Since HPLC-FTMS<sup>n</sup> only provided a MSI MI level 2 or 3 for the analyzed metabolites and relative quantifications, we applied SPE trapping of selected LC-MS peaks, based on their mass spectral characteristics, and subsequently

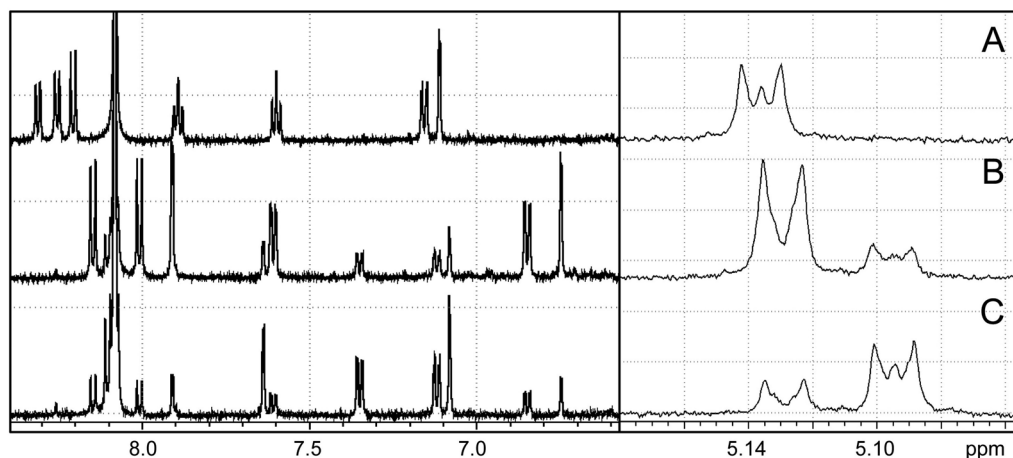
performed NMR measurements for a complete structural elucidation at MSI MI level 1.

*Metabolite identification using MS<sup>n</sup> and NMR.*

Structural elucidation of the SPE trapped conjugates was performed based on i) the HPLC-FTMS<sup>n</sup> data by comparing elution order, accurate masses, and fragments to those reported in literature<sup>21,22</sup> and ii) SPE-NMR data by comparison to available NMR data of phenolic core structures,<sup>39,44-46</sup> as well as iii) structural confirmation of metabolite conjugates by comparison of experimental and predicted NMR spectra of conjugates using the PERCH NMR software.<sup>38</sup> All metabolite annotations with their MSI metabolite identification level are listed in Supplemental Table 1.

*Small phenolic acid and urolithin conjugates in urine after black tea consumption:* From black tea-derived urine, 25 compounds were selected for NMR identification, based on their HPLC-FTMS<sup>n</sup> fragmentation patterns, including two vanillic acid glucuronides and three compounds that had accurate masses that were in agreement with xanthone or urolithin derivatives.<sup>47</sup> The two vanillic acid conjugates (m/z of 343.0670, C<sub>14</sub>H<sub>15</sub>O<sub>10</sub> [M-H]<sup>-</sup>) were present at significant amounts in the probe, i.e. 22 nmol (8.5 µgram) and 91 nmol (35 µgram) respectively. These analytes were identified based on 1D-<sup>1</sup>H-NMR and confirmative 2D-NMR experiments as vanillic acid-4-*O*-glucuronide and the vanillic acid glucuronidated on the carboxyl group, which was also confirmed by NMR predictions of the 1D-<sup>1</sup>H-NMR spectra. Sulphate and glucuronide conjugates of *p*-cresol (4-methylphenol) were also trapped in relative high amounts (i.e. 27 and 40 nmol respectively).

After trapping of three analytes having xanthone or urolithin-like HPLC-FTMS<sup>n</sup> spectral properties, we observed NMR signals with good signal to noise that we could assign to urolithin glucuronides (Figure 2). For one of the molecules, 2D-NMR (COSY, TOCSY, HSQC, HMBC) spectra were obtained and the structure of this molecule was annotated as urolithin-B-glucuronide (Figure 2A). We were also able to obtain a good fit of the 1D-<sup>1</sup>H-NMR data of this



**Figure 2:** Aromatic protons (6.4 – 8.4 ppm, left) and anomeric protons (5.06 – 5.16 ppm, right) in the NMR spectra of urolithin B glucuronide (A), urolithin A-8-O-glucuronide (B), and urolithin A-3-O-glucuronide (C). Spectra are obtained of SPE-trapped analytes of a urine fraction after black tea consumption. The typical pattern observed for the anomeric H1' proton around 5.10 ppm for the 3-O substituted urolithins can be rationalized by overlap of the H2' and H3' resonances.

urolithin-B-glucuronide. The other two urolithin analytes were proposed to be urolithin-A-glucuronide derivatives based on their accurate mass data and MS fragmentation patterns. Using the predicted spectral characteristics and by fitting of the 1D-<sup>1</sup>H-NMR datasets we could identify these two molecules as urolithin-A-8-O-glucuronide and urolithin-A-3-O-glucuronide. A characterizing feature of the 3-OH glucuronidation was the splitting pattern of the anomeric glucuronide protons (H1') as visible in figure 2A (5.13 ppm) and in figure 2C (5.08 ppm), which is indicative of strongly overlapping glucuronide H2' and H3' resonances. By fitting the mixture of the two urolithin-A isomers we were able to extract the precise chemical shifts of the glucuronide H2' and H3' protons. The H2' and H3' protons were separated by only 0.006 ppm for the 3-O-glucuronide whereas in the 8-O-glucuronide they were separated by 0.015 ppm. Further structural confirmation was achieved by observing the characteristic UV spectrum for 3,8-dihydroxylation of urolithin in the HPLC-DAD chromatogram and the fact that these two molecules co-eluted, in contrast to what was expected for 3,9-dihydroxylation with the use of our gradient.<sup>47</sup>

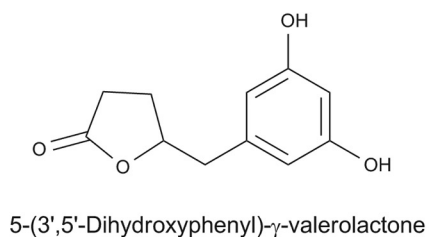
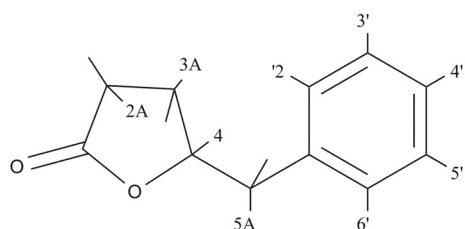
*Valerolactone conjugates in urine after green tea intake: HPLC-FTMS<sup>n</sup> profiling of urine extracts, after green tea consumption, yielded fragmentation data of 25 valerolactone conjugates, including several series of isomers, of which 11 were subsequently trapped for generating 1D-<sup>1</sup>H-NMR spectra. The valerolactone identity was confirmed by the 1D-<sup>1</sup>H-NMR spectra of trapped LC-TOFMS*

**Table 1:** *NMR shifts (in ppm) of aromatic protons of 11 conjugates that were analyzed using HPLC-TOFMS-SPE-NMR are plotted. The numbering is as denoted in the left structure below the table. The structure below at the right is an example of a valerolactone core structure, i.e. 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone. The proton NMR shifts are denoted with in brackets the difference between experimental and PERCH predicted value. Compound #5 is methylated on the 3'-OH position.*

ID	OH	O-Glc	O-SO <sub>3</sub>	H2'	H3'	H4'	H5'	H6'
1	3'	3'		7.005 (0.15)		7.01 (0.092)	7.26 (0.046)	6.97 (0.149)
2	4'	4'		7.226 (0.123)	7.062 (0.098)		7.062 (0.098)	7.226 (0.123)
3	3',4'	3'		7.071 (-0.002)			6.803 (-0.088)	6.84 (-0.204)
4	3',4'	4'		6.81 (0.145)			7.01 (-0.006)	6.72 (0.068)
5	3',4'	4'		6.96 (0.097)			7.01 (-0.053)	6.83 (0.047)
6	3',5'	3'		6.863 (-0.055)		6.427 (0.106)		6.105 (-0.082)
7	3',4',5'	3'		6.63 (0.058)				6.51 (0.301)
8	3',4',5'	4'		6.36 (0.079)				6.36 (0.079)

9	3',4'		3'	7.02 (-0.104)			7.01 (0.086)	6.95 (0.137)
10	3',5'		3'	6.71 (0.102)		6.69 (0.155)		6.52 (0.222)
11	3',4',5'		3'	6.71 (-0.013)				6.56 (0.196)

#	ID
1	5-(3'-hydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -glucuronide
2	5-(4'-hydroxyphenyl)- $\gamma$ -valerolactone-4'- <i>O</i> -glucuronide
3	5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -glucuronide
4	5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-4'- <i>O</i> -glucuronide
5	5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -methyl-4'- <i>O</i> -glucuronide
6	5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -glucuronide
7	5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -glucuronide
8	5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-4'- <i>O</i> -glucuronide
9	5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -sulphate
10	5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -sulphate
11	5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'- <i>O</i> -sulphate





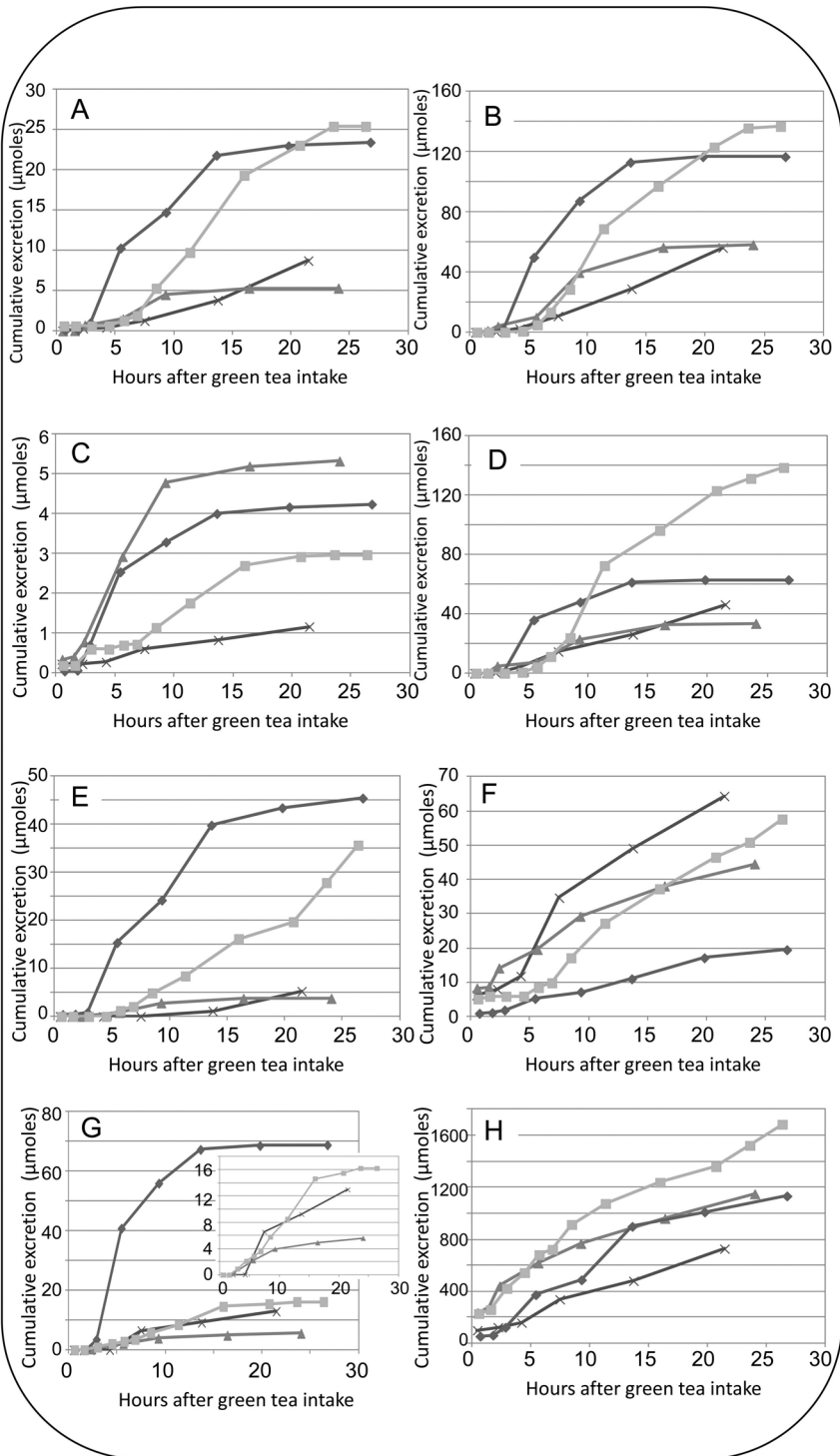
peaks showing two double doublets of H5A and H5B with coupling constants ( $J$ 's) of 13 and 6 Hz between 3.1 and 2.7 ppm, respectively.<sup>39, 44</sup> Table 1 presents the experimental aromatic NMR shifts of the valerolactone conjugates and the differences with the PERCH NMR software predictions. The chemical shifts of at least one of the aromatic signals of the conjugated valerolactones showed large differences with experimental and predicted NMR shifts for the corresponding non-conjugated valerolactone.<sup>39, 44</sup> We were able to identify the protons for which the chemical shifts were strongest influenced by the conjugate group by their predicted shift change and coupling constants. All differences between experimental and predicted chemical shifts of the protons close to the conjugate group were within 0.155 ppm. Larger differences were obtained for the protons (as for example H6') that were further away of the conjugate group. This is most probably due too long range interactions that are not yet well represented in the prediction model. We were able to confirm the NMR peak assignments of 7 valerolactones for which the difference between predicted and experimental chemical shifts was within 0.155 ppm. Especially in case of partially pure analytes, the PERCH analysis helped in assigning the correct NMR peaks, as for example for 5-(4'-hydroxyphenyl)- $\gamma$ -valerolactone-4'- $O$ -glucuronide. The aromatic peaks confirmed phenolic ring substitution, i.e. for 3',4'-dihydroxylation the H2' appeared as small doublet ( $J$  of 2 Hz), H5' as doublet ( $J$  of 8 Hz) and H6' as double doublet ( $J$ 's of 8 and 2 Hz), and the peak area integrals always matched with those of H5A and H5B. Compared to the non-conjugated 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone, the chemical shifts of the aromatic H5' protons were shifted by 0.320 ppm upon 4'- $O$ -glucuronidation whereas the chemical shifts of the other two aromatic protons were almost unchanged. The strong change in the ppm values for the H5' was also well predicted by the PERCH NMR software. Altogether, we could structurally elucidate 10 monoconjugated valerolactones up to MSI MI level 1 and one diconjugated valerolactone at MSI MI level 2. The SPE trapped amounts in the NMR probe ranged from 74 nmol (32  $\mu$ gram) for 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'- $O$ -glucuronide down to 1.6 nmol (0.5  $\mu$ gram) for 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'- $O$ -glucuronide

based on quantification with rutin as a standard. The 5-(3',5'-hydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-sulphate was partly deconjugated after SPE trapping. We observed the proton signals belonging to the sulphated conjugate in the 1D-<sup>1</sup>H-NMR spectrum by an additional set of H5A and H5B proton signals that were shifted upfield compared to the deconjugated valerolactone signals, as well as aromatic peaks that matched in terms of their NMR peak area integrals.

*NMR based quantification of urinary excretion.*

Diagnostic proton signals in the 1D-<sup>1</sup>H-NMR spectra of the SPE trapped analytes were used to quantify urinary excretion of identified conjugates after green tea intake in the 1D-<sup>1</sup>H-NMR spectra of urine fractions (Supplemental Figure S3). Figures 3A-3H represent the cumulative excretion of urinary conjugates in  $\mu$ moles over the 26 hours sampling time for the 4 volunteers. The total cumulative excretion of urinary conjugates ranged from 1 - 5  $\mu$ moles for 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-4'-*O*-glucuronide to 700 - 1600  $\mu$ moles for hippuric acid. Significant differences in cumulative excretion profiles of the isomers of 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone glucuronide were observed in the volunteers, i.e., the 3'-*O*-glucuronide was excreted about 5 - 20 times more than the 4'-*O*-glucuronide conjugate. To compare these 1D-<sup>1</sup>H-NMR based quantitative patterns with LC-MS based relative quantifications of phenol conjugates in urine after green tea intake, we plotted the cumulative excretion based on NMR against the cumulative excretion based on relative abundance determined from HPLC-FTMS peak areas for five compounds present in the 4

**Figure 3 (on next page):** NMR based cumulative urinary excretion over 26 hours of 6 unique conjugates and 2 mixed conjugates (signals almost overlap) in 4 volunteers (volunteer 1 [ $\times$ ], 2 [ $\blacktriangle$ ], 3 [ $\blacklozenge$ ], and 4 [ $\blacksquare$ ]), plotted as micromoles in urine fraction after green tea intake, i.e. 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-glucuronide (A), 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-glucuronide (B), 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-4'-*O*-glucuronide (C), 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-sulphate and its 4-hydroxyvaleric acid analogue (D), 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-sulphate (E), indole-3-acetic-acid-*O*-glucuronide (F), pyrogallol-2-*O*-sulphate and pyrogallol-2-*O*-glucuronide (G), and hippuric acid (H).



volunteers (Suppl. Fig. S4). Apart from two outlier points, a significant linear correlation ( $R^2$  of 0.88) was observed between the two analytical platforms.

### *Discussion*

In this study we present an efficient approach to obtain full structural identification and quantification of phenolic conjugates present in human urine after tea consumption. To overcome current obstacles in metabolite identification of conjugates present at micromolar concentrations, we combined a SPE concentration and purification step with the combined use of HPLC-FTMS<sup>n</sup> and HPLC-TOFMS-SPE-NMR to obtain sufficient analytical evidence for structural elucidation of urinary conjugates present at micromolar levels.

The SPE purification and concentration resulted in a less complex background, for example due to removal of salt and urea and background of polar metabolites, and at the same time concentrated the urine fractions 15 fold, which facilitated the acquisition of NMR spectra containing phenolic conjugate signals with sufficient signal to noise ratio. The HPLC-FTMS<sup>n</sup> method was used to detect and select phenolic conjugates based on their accurate mass and mass spectral fragmentations.<sup>25</sup> The elucidation of a series of valerolactone conjugates shows that candidate restriction by using HPLC-FTMS<sup>n</sup> is effective, because only a few possible structures remained valid for consideration during the interpretation of NMR data. Subsequently, the use of HPLC-TOFMS-SPE trapping separated the NMR signals of structurally related molecules, which provided the complementary structural information needed for complete elucidation. Additionally, 1D-<sup>1</sup>H-NMR spectra confirmed core identities, for example of valerolactones by the specific double doublet of the H5 protons.

Most SPE trapped LC-MS peaks contained sufficient analyte for 1D-<sup>1</sup>H-NMR experiments, but not sufficient for elaborate and time consuming 2D-NMR experiments. Therefore, NMR peak integrations and spectral predictions with the PERCH NMR software were very helpful in confirming the correct assignments of NMR peaks, especially in case of partially pure NMR spectra. The combination of MS and NMR based analytical platforms in compound identification has been

applied before.<sup>26, 48, 49</sup> The complementarity of MS and NMR is shown in the structural elucidation of NMR silent conjugated groups as sulphates, as neither of the analytical techniques provides unambiguous identification.<sup>50, 51</sup> Here, we extend this analytical tandem by using accurate mass MS<sup>n</sup> data that provide a wealth of structural information for metabolite identification. Altogether, our analytical approach resulted in the elucidation of 36 phenolic conjugates (MSI MI level 1) and in total annotated 138 metabolites (MSI MI level 2 and 3) from human urine using the combination of HPLC-FTMS<sup>n</sup> and TOFMS-guided HPLC-SPE-NMR. Finally, the (partially) purified analytes could be used to quantify a number of identified metabolites, using their specific 1D-<sup>1</sup>H-NMR signals, without the need for purifying large amounts as reference compounds.

*Identification of conjugates of microbial bioconversion products of tea polyphenols:* Metabolite identification is still a major issue in nutritional metabolomics.<sup>52</sup> Recently published studies confirm the distribution of conjugated and non-conjugated flavan-3-ols in body tissues<sup>53</sup> and underpin the importance of the complete identification of dietary polyphenols and their metabolized conjugates, as their potential biological functions as protein ligands and affinity for enzymes is dependent on their exact 3D structure.<sup>54-56</sup> Also, the complete identification of 36 metabolites complements dedicated databases of the food metabolome that are currently being set up.<sup>57</sup> In the identification process of metabolite conjugations, the PERCH NMR software was useful in assigning the conjugation site on the phenolic core and we foresee that in future it will become more incorporated in NMR based structural elucidation.<sup>58</sup>

*Metabolite quantification of urine conjugates:* NMR is a quantitative approach.<sup>59</sup> However, metabolite quantification is still a challenge in the complex 1D-<sup>1</sup>H-NMR spectra such as derived from urine. Overlap of diagnostic NMR proton signals with other NMR signals, for example, limit their use in quantification. The use of SPE fractionation allows for quantitative information on the excretion of urinary conjugates without the need for (non-existing) reference compounds. The total urinary excretion based on NMR and the excretion based on LC-MS showed a good correlation (Supplemental Figure S4). Detailed analysis of the

NMR spectra revealed that the two outliers were the result of overlapping peaks in the region used for quantification. In the future, we foresee that conjugate quantification could be performed in an automatic manner by deconvolution of (partially purified) 1D-<sup>1</sup>H-NMR spectra using the PERCH NMR software.<sup>38, 60, 61</sup>

*Aromatic urinary metabolites excreted in urine after black and green tea intake:* Hippuric acid and a structurally related hydroxybenzoic glycine conjugate, vanilloylglycine, and pyrogallol-2-*O*-sulphate were observed in urine after both black and green tea consumption based on their NMR as well as HPLC-FTMS<sup>n</sup> data (Fig. 3, Suppl. Table S1). Hippuric acid and pyrogallol-2-*O*-sulphate were earlier observed in urine after black tea intake.<sup>2, 12, 34</sup> In fact, hippuric acid is the end product of several breakdown pathways, like those for dietary phenolics and aromatic amino acids, and its concentration is also dependent on the basal diet,<sup>52</sup> and hippuric acid is therefore not likely to be a good marker of tea intake.

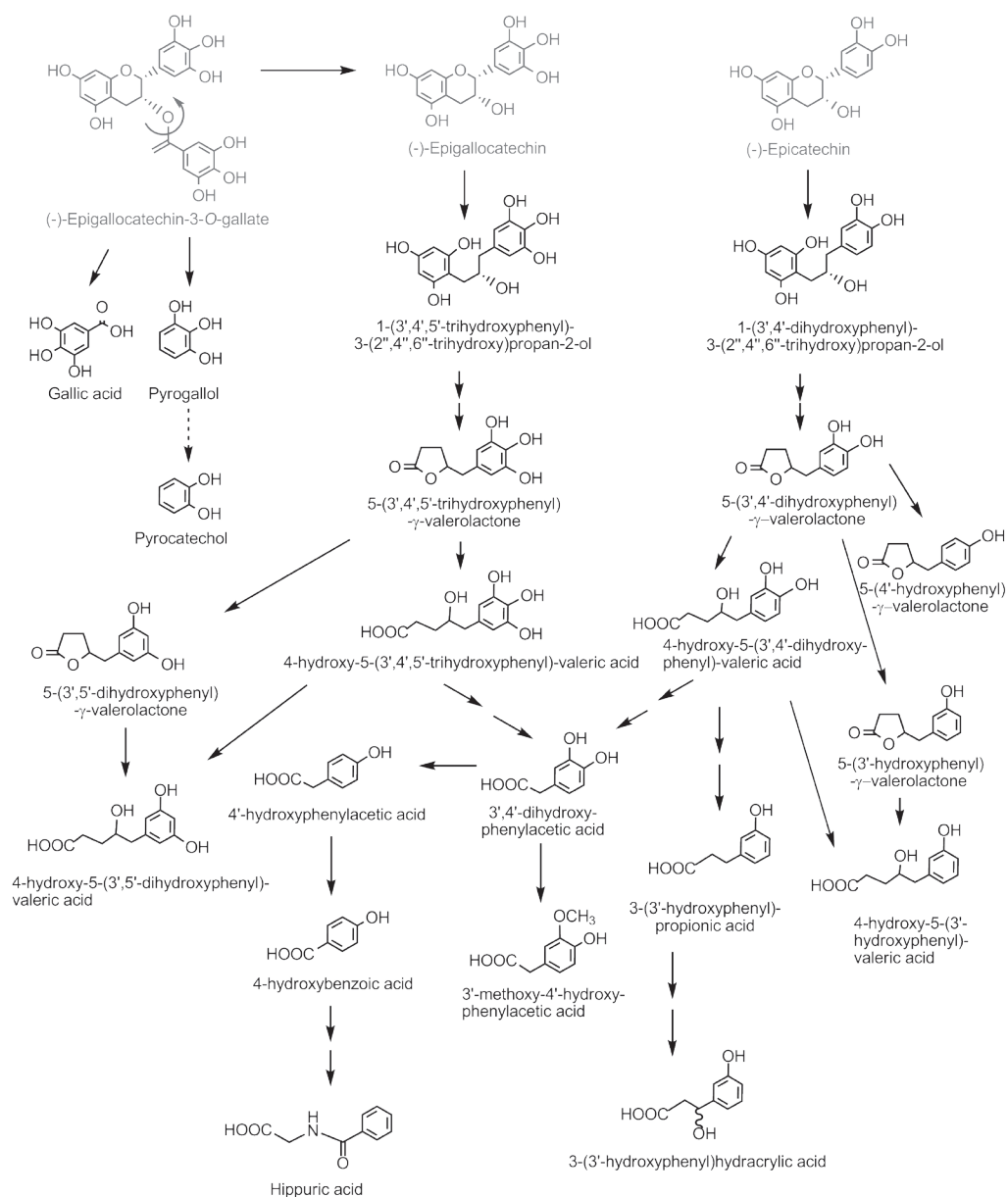
*Metabolic fate of green tea flavan-3-ols.*

*HPLC-FTMS profiles of conjugated polyphenols:* Numerous studies describe a low absorption of polyphenol glycosides by the small intestine, with the largest part of polyphenols arriving in the large intestine, where microbes break them down.<sup>2, 32</sup> Nevertheless, we and others detected (epi)catechin and (epi)gallocatechin conjugates in urine upon green tea consumption<sup>20</sup> (Supplemental Table 1). The extracted ion HPLC-FTMS chromatograms we obtained for sulphated and methylated (epi)catechin and (epi)gallocatechin are relatively similar for the 4 volunteers. Actually, full scan LC-MS profiles, as plotted for the (epi)catechin-*O*-sulphate-*O*-methyl isomers (*m/z* 383.0444, with the EF C16H15O9S [M-H]<sup>-</sup>) in Suppl. Fig. 5, were remarkably similar to extracted ion chromatogram profiles in plasma reported by Stallmach et al.<sup>20</sup> The low variation between volunteers in the LC-MS profiles of metabolites absorbed in the small intestine suggest that these profiles could be valuable short-term markers of intake for polyphenols.

*Valerolactone excretion:* Free valerolactones and conjugates thereof have been detected in urine, plasma, and faecal slurries after intake of various polyphenol

containing food and beverages.<sup>21, 22, 44, 62</sup> However, until now, no complete identification and quantification of conjugates in urine could be achieved. Sánchez-Patan et al. found similar metabolites in their free forms in *in vitro* batch models with faecal slurries of three volunteers incubated with grape seed extract and the authors identified 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone as abundant microbial breakdown product.<sup>62, 63</sup> Based on their data, enzymatic dehydroxylation preferably takes place at the 4'-O position resulting in 3'-hydroxyl conjugated phenolic metabolites. This is reflected in our data by the quantitative difference between 10 nmol 5-(3'-hydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-conjugate and 4 nmol of its 4'-analogue in the NMR probe after SPE trapping, which was similar to the relative differences observed in HPLC-FTMS areas. It should be noted here that the different valerolactones may have (slightly) different recoveries during the SPE-NMR trapping, due to the different amounts of acetonitrile as a consequence of the LC gradient. Unfortunately, it was not feasible to directly quantify the 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone conjugates in the SPE concentrated urine fractions, due to signal overlap in the complex 1D-<sup>1</sup>H-NMR spectra. Nevertheless, our SPE-NMR data suggests that 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-*O*-glucuronide was quantitatively the most abundant valerolactone conjugate in urine (74 nmoles present in the NMR probe after SPE trapping), about a factor 7 more than the two second-most abundant conjugates. It has been hypothesized that 3',4'-dihydroxy-phenylvalerolactone conjugates are derived of (epi)catechin, whereas 3',4',5'-trihydroxy-phenylvalerolactone conjugates are formed from (epi)gallocatechin, which is also true for the galloylated flavan-3-ol forms<sup>10</sup> (Figure 4). Using the relation between the NMR based and HPLC-FTMS<sup>n</sup> based quantification (Suppl. Fig 4), our data suggests that the (combined) 3',4'-dihydroxy-phenylvalerolactone conjugates are a factor 12 to 20 more abundant than the (combined) 3',5'-dihydroxy-phenylvalerolactone and 3',4',5'-trihydroxy-phenylvalerolactone conjugates. Interestingly, (epi)gallocatechin metabolites are the more abundant flavan-3-ols present in green tea.<sup>28</sup> Thus, our data suggests a preference of gut microbes to metabolize





**Figure 4:** Schematic representation of the metabolism of flavan-3-ols in the human body is shown. The structures are found in human urine as conjugates with glucuronide, sulphate, and/or methyl groups. Scheme was adapted from Roowi et al.,<sup>10</sup> and complemented with results from the present study.



3',4'-dihydroxyl substituted flavan-3-ols into valerolactones. In addition, we could determine that the 3'-OH group is the preferred glucuronidation position over the 4'-OH group for the 3',4'-dihydroxy-phenylvalerolactone conjugates by a factor ranging from 2.7 to 3.7, while the sulphated conjugates showed a higher preference for the 3'-OH group over the 4'-OH group with a larger variation ranging from 6 to 41.

All these observations show high inter-subject differences in nutrkinetics, as can be seen in Figure 3 as well. For example, in plots 3A and 3B volunteers 3 and 4 are likely to be high responders, while volunteers 1 and 2 would be low responders.<sup>42</sup> Similarly, volunteer 3 is then a fast responder, while volunteer 1 is a slow responder.<sup>42</sup> As the scope of this study was the structural elucidation of metabolite conjugates and their subsequent quantification, a placebo intake was not performed. However, the levels of baseline excretion of valerolactones are expected to be neglectable when compared to the excretion as a result of polyphenol intake, while for hippuric acid a baseline correction would be needed to correctly assess the contribution of phenolic compounds to its excretion (<sup>42</sup>, and Figure 3H). There were also some general characteristics observed in the plots: the total cumulative excretion of all valerolactone and valeric acid metabolites was reached between 5 and 10 hours, indicating that these metabolites are produced by gut microbes. To compare urinary metabolites present and their cumulative excretion profiles between black or green tea intake, a more elaborate study is needed. In this study, the 4 urolithin conjugates were only present after black tea intake.

Using the 48 HPLC-FTMS<sup>n</sup> annotated valerolactone and valeric acid conjugates we extended the present microbial breakdown pathway of flavan-3-ols in humans<sup>10</sup> (Figure 4). Single and double conjugated valerolactone and valeric acid metabolites, as well as smaller phenolic acids were annotated, and especially a range of methylated conjugates were annotated that have not been reported before. Numerous conjugates of the small phenolic acids, for example two sulphonated methylgallic acids, were annotated based on HPLC-FTMS<sup>n</sup> spectra (Supplemental Table 1).

*Identification of urolithin and indole glucuronides.*

Urolithins are known to be microbial breakdown products of ellagic acid, a compound that has been reported to be present in strawberries, raspberries, pomegranates, walnuts, and tea.<sup>31, 47</sup> We were able to obtain good quality 1D-<sup>1</sup>H-NMR spectra of glucuronidated conjugates (Figure 2). Surprisingly we observed similar quantities of urolithin-B and urolithin-A molecules which is in contrast to previous studies.<sup>64</sup> Moreover, we observed 8-O glucuronidation of urolithin A, in contrast to González-Barrio et al., who reported 3-O and 9-O glucuronidation.<sup>47</sup> The two urolithin-A glucuronides coeluted and showed similar fragmentation patterns in HPLC-FTMS<sup>n</sup>. We were able to completely characterize the 8-O glucuronide only by fitting the 1D-<sup>1</sup>H-NMR spectra of the mixture of the two isomers. The precise predictions of the glucuronide H2' and H3' protons confirmed structural features like 3-O glucuronidation and the distinct UV characteristics of 3,8 and 3,9 dihydroxylation excluded the structure of isourolithin A. This underlines the need to complement MS with NMR, and if needed DAD, for improved profiling of urine extracts. Indole-3-acetic acid (IAA) and indole-3-carboxylic acid glucuronides were found in to be present in urine in substantial quantities upon green and black tea intake. IAA has been reported to be present in human urine as an indolic metabolite from tryptophan;<sup>65</sup> however, the identified indolic conjugates were not completely resolved before. The fact that all four volunteers showed a substantial excretion of these indolic compounds after green tea intake may indicate that the consumption of green tea could influence human biochemical pathways related to tryptophan.

In conclusion, using the combination of HPLC-FTMS<sup>n</sup> and 1D-<sup>1</sup>H-NMR holds great promise for future work in food and nutritional metabolomics and related fields.<sup>7, 66, 67</sup> Our analytical strategy can efficiently identify both food-derived exogenous metabolites after tea intake as well as endogenous metabolites present in urine, and monitor their levels in a quantitative manner. The HPLC-FTMS<sup>n</sup> method rapidly provided insight in the presence of molecules related to

intake of polyphenols, and the coupling with SPE-NMR enabled the efficient elucidation and confirmation of their identity. Identified structures are the basis to gain further understanding of the influence of the human host on the so-called food metabolome. In addition, the identified and partly or completely purified analytes can be used for quantification purposes in future urine or even plasma studies, for example as HPLC-MS standards or for the creation of (HPLC-)FTMS<sup>n</sup> databases for (xenobiotic) urine and plasma metabolites.<sup>68,69</sup> This purification of conjugates present in human body fluids followed by NMR measurements is an alternative to organic synthetic approaches that produced reference compounds for quantification.<sup>39, 63, 70</sup> Additionally, metabolites could be used in bioactivity assays to provide insight in their mechanism of action.<sup>52, 68</sup> Therefore, we believe that the combination of HPLC-MS<sup>n</sup> and SPE-NMR greatly contributes to progress in nutritional metabolomics and polyphenol metabolic fate studies in particular.

### *Acknowledgements*

This research was granted by The Netherlands Metabolomics Centre and The Centre for BioSystems Genomics, both of which are part of The Netherlands Genomics Initiative/Netherlands Organization for Scientific Research. The authors thank dr. Noriyuki Nakajima, Monique Klinkenberg, Moktar Akermi, and Jonathan Bodin for their contribution.

### *Supporting Information*

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>. The Supporting Information is also part of this thesis, pages 197-206, and the Excel file is also available at <http://edepot.wur.nl/216854>

## References

1. J. P. E. Spencer, M. M. A. E. Mohsen, A.-M. Miniñane and J. C. Mathers, *Brit. J. Nutr.*, 2008, **99**, 12-22.
2. J. van Duynhoven, E. E. Vaughan, D. M. Jacobs, R. A. Kemperman, E. J. J. van Velzen, G. Gross, L. C. Roger, S. Possemiers, A. K. Smilde, J. Doré, J. A. Westerhuis and T. Van de Wiele, *P.N.A.S.*, 2010, **108**, 4531-4538.
3. H. Schroeter, C. Heiss, J. P. E. Spencer, C. L. Keen, J. R. Lupton and H. H. Schmitz, *Mol. Asp. Med.*, 2010, **31**, 546-557.
4. D. Del Rio, L. G. Costa, M. E. J. Lean and A. Crozier, *Nutr. Metab. Cardiovasc. Dis.*, 2011, **20**, 1-6.
5. E. Verzelloni, C. Pellacani, D. Tagliazucchi, S. Tagliaferri, L. Calani, L. G. Costa, F. Brighenti, G. Borges, A. Crozier, A. Conte and D. Del Rio, *Mol. Nutr. Food Res.*, 2011, **55**, S35-S43.
6. M. Monagas, M. Urpi-Sarda, F. Sánchez-Patán, R. Llorach, I. Garrido, C. Gómez-Cordovés, C. Andres-Lacueva and B. Bartolomé, *Food Funct.*, 2010, **1**, 233-253.
7. C. Manach, J. Hubert, R. Llorach and A. Scalbert, *Mol. Nutr. Food Res.*, 2009, **53**, 1303-1315.
8. S. Sang, J. D. Lambert, C.-T. Ho and C. S. Yang, *Pharmacol. Res.*, 2011, **64**, 87-99.
9. D. Del Rio, A. Stalmach, L. Calani and A. Crozier, *Nutrients*, 2011, **2**, 820-833.
10. S. Roowi, A. Stalmach, W. Mullen, M. E. J. Lean, C. A. Edwards And and A. Crozier, *J. Agr. Food Chem.*, 2010, **58**, 1296-1304.
11. W. Y. Feng, *Curr. Drug Metab.*, 2006, **7**, 755-809.
12. C. A. Daykin, J. P. M. Van Duynhoven, A. Groenewegen, M. Dachtler, J. M. M. Van Amelsvoort and T. P. J. Mulder, *J. Agr. Food Chem.*, 2005, **53**, 1428-1434.
13. M.-J. Lee, P. Maliakal, L. Chen, X. Meng, F. Y. Bondoc, S. Prabhu, G. Lambert, S. Mohr and C. S. Yang, *Canc. Epidemiol. Biomark. Prevent.*, 2002, **11**, 1025-1032.
14. M. Urpi-Sarda, M. Monagas, N. Khan, R. Llorach, R. M. Lamuela-Rotchés, O. Jáuregui, R. Estruch, M. Izquierdo-Pulido and C. Andrés-Lacueva, *J. Chromatogr. A*, 2009, **1216**, 7258-7267.
15. H. Ito, M. P. Gonthier, C. Manach, C. Morand, L. Mennen, C. Rémésy and A. Scalbert, *Brit. J. Nutr.*, 2005, **94**, 500-509.
16. M. Renouf, K. Redeuil, K. Longet, C. Marmet, F. Dionisi, M. Kussmann, G. Williamson and K. Nagy, *Eur. J. Nutr.*, 2011, **50**, 575-580.
17. T. Farrell, L. Poquet, F. Dionisi, D. Barron and G. Williamson, *J. Pharmaceut. Biomed. Anal.*, 2011, **55**, 1245-1254.
18. E. Roura, C. Andres-Lacueva, O. Jauregui, E. Badia, R. Estruch, M. Izquierdo-Pulido and R. M. Lamuela-Raventos, *J. Agr. Food Chem.*, 2005, **53**, 6190-6194.
19. A. K. Savage, J. P. M. Van Duynhoven, G. Tucker and C. A. Daykin, *Magn. Res. Chem.*, 2011, **49**, S27-S36.

20. A. Stalmach, W. Mullen, H. Steiling, G. Williamson, M. E. J. Lean and A. Crozier, *Mol. Nutr. Food Res.*, 2010, **54**, 323-334.
21. S. Sang and C. S. Yang, *Rapid Commun. Mass Spectrom.*, 2008, **22**, 3693-3699.
22. R. Llorach, I. Garrido, M. Monagas, M. Urpi-Sarda, S. Tulipani, B. Bartolome and C. Andres-Lacueva, *J. Prot. Res.*, 2010, **9**, 5859-5867.
23. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, **2**, 23-60.
24. D. S. Wishart, *Bioanalysis*, 2011, **3**, 1769-1782.
25. J. J. J. Van der Hooft, J. Vervoort, R. J. Bino and R. C. H. de Vos, *Metabolomics*, 2012, **8**, 691-703.
26. S. Moco, J. Vervoort, R. J. Bino, R. C. H. De Vos and R. Bino, *Tr. Anal. Chem.*, 2007, **26**, 855-866.
27. J. J. J. Van der Hooft, V. Mihaleva, R. J. Bino, R. C. H. de Vos and J. Vervoort, *Magn. Res. Chem.*, 2011, **49**, S55-S60.
28. J. Peterson, J. Dwyer, S. Bhagwat, D. Haytowitz, J. Holden, A. L. Eldridge, G. Beecher and J. Aladesanmi, *J. Food Comp. Anal.*, 2005, **18**, 487-501.
29. L.-Z. Lin, P. Chen and J. M. Harnly, *J. Agr. Food Chem.*, 2008, **56**, 8130-8140.
30. S. Scharbert, N. Holzmann and T. Hofmann, *J. Agr. Food Chem.*, 2004, **52**, 3498-3508.
31. J. J. J. van der Hooft, M. Akermi, F. Y. Ünlü, V. Mihaleva, V. G. Roldan, R. J. Bino, R. C. H. De Vos and J. Vervoort, *J. Agr. Food Chem.*, 2012, **Article ASAP - DOI 10.1021/jf300297y**.
32. C. Manach and J. L. Donovan, *Free Rad. Res.*, 2004, **38**, 771-785.
33. P. C. H. Hollman, M. N. C. P. Bijman, Y. van Gameren, E. P. J. Cnossen, J. H. M. de Vries and M. B. Katan, *Free Rad. Res.*, 1999, **31**, 569-573.
34. F. A. Van Dorsten, C. A. Daykin, T. P. J. Mulder and J. P. M. Van Duynhoven, *J. Agr. Food Chem.*, 2006, **54**, 6929-6938.
35. E. J. J. Van Velzen, J. A. Westerhuis, J. P. M. Van Duynhoven, F. A. Van Dorsten, C. H. Grön, D. M. Jacobs, G. S. M. J. E. Duchateau, D. J. Vis and A. K. Smilde, *J. Prot. Res.*, 2009, **8**, 3317-3330.
36. J. J. J. Van der Hooft, J. Vervoort, R. J. Bino, J. Beekwilder and R. C. H. De Vos, *Anal. Chem.*, 2011, **83**, 409-416.
37. R. Laatikainen, M. Niemitz, U. Weber, J. Sundeun, T. Hassinen and J. Vepsäläinen, *J. Magn. Res. Ser. A*, 1996, **120**, 1-10.
38. PERCH, <http://www.perchsolutions.com>, 2011.
39. M. Hamada, A. Furuno, S. Nakano, T. Kishimoto and N. Nakajima, *Synthesis*, 2010, 1512-1520.
40. G. Gross, D. M. Jacobs, S. Peters, S. Possemiers, J. van Duynhoven, E. E. Vaughan and T. van de Wiele, *J. Agr. Food Chem.*, 2010, **58**, 10236-10246.

41. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, **3**, 211-221.
42. J. P. M. van Duynhoven, E. J. J. van Velzen, J. A. Westerhuis, M. Foltz, D. M. Jacobs and A. K. Smilde, *Tr. Food Sci. Technol.*, 2012, **26**, 4-13.
43. G. Borges, W. Mullen, A. Mullan, M. E. J. Lean, S. A. Roberts and A. Crozier, *Mol. Nutr. Food Res.*, 2010, **54**, S268-S277.
44. A. Takagaki and F. Nanjo, *J. Agr. Food Chem.*, 2010, **58**, 1313-1321.
45. T. Kohri, N. Matsumoto, M. Yamakawa, M. Suzuki, F. Nanjo, Y. Hara and N. Oku, *J. Agr. Food Chem.*, 2001, **49**, 4102-4112.
46. T. Kohri, M. Suzuki and F. Nanjo, *J. Agr. Food Chem.*, 2003, **51**, 5561-5566.
47. R. González-Barrio, P. Truchado, H. Ito, J. C. Espín and F. A. Tomás-Barberán, *J. Agr. Food Chem.*, 2011, **59**, 1152-1162.
48. W. S. Law, P. Y. Huang, E. S. Ong, C. N. Ong, S. F. Y. Li, K. K. Pasikanti and E. C. Y. Chan, *Rapid Commun. Mass Spectrom.*, 2008, **22**, 2436-2446.
49. J. L. Wolfender, G. Marti and E. F. Queiroz, *Current Organic Chemistry*, 2010, **14**, 1808-1832.
50. J. P. Shockcor, S. E. Unger, I. D. Wilson, P. J. D. Foxall, J. K. Nicholson and J. C. Lindon, *Analytical Chemistry*, 1996, **68**, 4431-4435.
51. U. G. Sidelmann, I. Bjørnsdottir, J. P. Shockcor, S. H. Hansen, J. C. Lindon and J. K. Nicholson, *J. Pharmaceut. Biomed. Anal.*, 2001, **24**, 569-579.
52. T. K. McGhie and D. D. Rowan, *Mol. Nutr. Food Res.*, 2012, **56**, 147-158.
53. M. Urpi-Sarda, E. Ramiro-Puig, N. Khan, S. Ramos-Romero, R. Llorach, M. Castell, S. Gonzalez-Manzano, C. Santos-Buelga and C. Andres-Lacueva, *Brit. J. Nutr.*, 2011, **103**, 1393-1397.
54. W. Brand, M. G. Boersma, H. Bik, E. F. Hoek-van Den Hil, J. Vervoort, D. Barron, W. Meinel, H. Glatt, G. Williamson, P. J. Van Bladeren and I. M. C. M. Rietjens, *Drug Metabol. Dispos.*, 2010, **38**, 617-625.
55. W. Brand, B. Oosterhuis, P. Krajcsi, D. Barron, F. Dionisi, P. J. Van Bladeren, I. M. C. M. Rietjens and G. Williamson, *Biopharmac. Drug Dispos.*, 2011, **32**, 530-535.
56. J. I. Ottaviani, T. Y. Momma, C. Heiss, C. Kwik-Urbe, H. Schroeter and C. L. Keen, *Free Rad. Biol. Med.*, 2011, **50**, 237-244.
57. A. Scalbert, C. Andres-Lacueva, M. Arita, P. Kroon, C. Manach, M. Urpi-Sarda and D. Wishart, *J. Agr. Food Chem.*, 2011, **59**, 4331-4348.
58. S. Moco, L. H. Tseng, M. Spraul, Z. Chen and J. Vervoort, *Chromatographia*, 2006, **64**, 503-508.
59. D. S. Wishart, *Tr. Anal. Chem.*, 2008, **27**, 228-237.

60. J. G. Napolitano, T. Gödecke, M. F. Rodríguez-Brasco, B. U. Jaki, S. N. Chen, D. C. Lankin and G. F. Pauli, *J. Nat. Prod.*, 2012, **75**, 238-248.
61. N. Aranibar, M. Borys, N. A. Mackin, V. Ly, N. Abu-Absi, S. Abu-Absi, M. Niemitz, B. Schilling, Z. J. Li, B. Brock, R. J. Russell II, A. Tymiak and M. D. Reily, *J. Biomol. NMR*, 2011, **49**, 195-206.
62. F. Sánchez-Patán, C. Cueva, M. Monagas, G. E. Walton, G. R. Gibson, P. J. Martín-Álvarez, M. Victoria Moreno-Arribas and B. a. Bartolomé, *Food Chem.*, 2012, **131**, 337-347.
63. F. Sánchez-Patán, M. Chioua, I. Garrido, C. Cueva, A. Samadi, J. Marco-Contelles, M. V. Moreno-Arribas, B. Bartolomé and M. Monagas, *J. Agr. Food Chem.*, 2011, **59**, 7083-7091.
64. P. Truchado, M. Larrosa, M. Teresa, García-Conesa, B. Cerdá, M. Luisa, Vidal-Guevara, F. A. Tomás-Barberán and J. C. Espín, *J. Agr. Food Chem.*, 2012, **60**, 5749-5754.
65. H. Weissbach, W. King, A. Sjoerdsma and S. Udenfriend, *J. Biol. Chem.*, 1959, **234**, 81-86.
66. A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart and S. Wopereis, *Metabolomics*, 2009, 1-24.
67. S. Primrose, J. Draper, R. Elsom, V. Kirkpatrick, J. C. Mathers, C. Seal, M. Beckmann, S. Haldar, J. H. Beattie, J. K. Lodge, M. Jenab, H. Keun and A. Scalbert, *Brit. J. Nutr.*, 2011, **105**, 1277-1283.
68. A. Mutlib, R. Espina, K. Vishwanathan, K. Babalola, Z. Chen, C. Dehnhardt, A. Venkatesan, T. Mansour, I. Chaudhary, R. Talaat and J. A. Scatina, *Drug Metabol. Dispos.*, 2011, **39**, 106-116.
69. G. S. Walker, T. F. Ryder, R. Sharma, E. B. Smith and A. Freund, *Drug Metabol. Dispos.*, 2011, **39**, 433-440.
70. E. S. Mull, M. Van Zandt, A. Golebiowski, R. P. Beckett, P. K. Sharma and H. Schroeter, *Tetrahedron Let.*, 2012, **53**, 1501-1503.





Supporting information to *Chapter 6*:  
‘Structural elucidation and quantification of  
phenolic conjugates present in human urine  
after tea intake’



Justin J.J. van der Hooft, Ric C.H. de Vos, Raoul J. Bino,  
Velitchka Mihaleva, Lars Ridder, Niels de Roo, Doris M. Jacobs,  
John P.M. van Duynhoven, and Jacques Vervoort

This Supporting Information was published online with the research article in *Analytical Chemistry*, 2012, volume 84 (16), pp 7263-7271, DOI: 10.1021/ac3017339

*Supplemental text 1: Experimental details*

Sample preparation for  $1D\text{-}^1H\text{-NMR}$ ,  $HPLC\text{-FTMS}^n$ ,  $HPLC\text{-TOFMS}\text{-SPE}\text{-NMR}$ , and  $Offline\text{-FTMS}^n$  measurements:

*$1D\text{-}^1H\text{-NMR}$  sample preparation:* The freeze dried samples were redissolved in 200  $\mu\text{l}$  of deuterated methanol upon 2 min sonication. Subsequently, the samples were transferred in 3 mm NMR tubes (Bruker Match system) that were capped and stored at 4°C before NMR analysis.

*$HPLC\text{-FTMS}^n$  sample preparation:* The freeze dried samples were redissolved in 200  $\mu\text{l}$  25% methanol in  $\text{H}_2\text{O}$  (urine samples after green tea consumption) or 75% MeOH in  $\text{H}_2\text{O}$  (urine after black tea consumption) containing 0.1% formic acid upon 5 minutes sonication and filtered through a 0.45  $\mu\text{m}$  filter before transferring in a HPLC vial. The vials were stored at -20 °C protected from light before  $HPLC\text{-FTMS}^n$  analysis. The total concentration factor was 15.

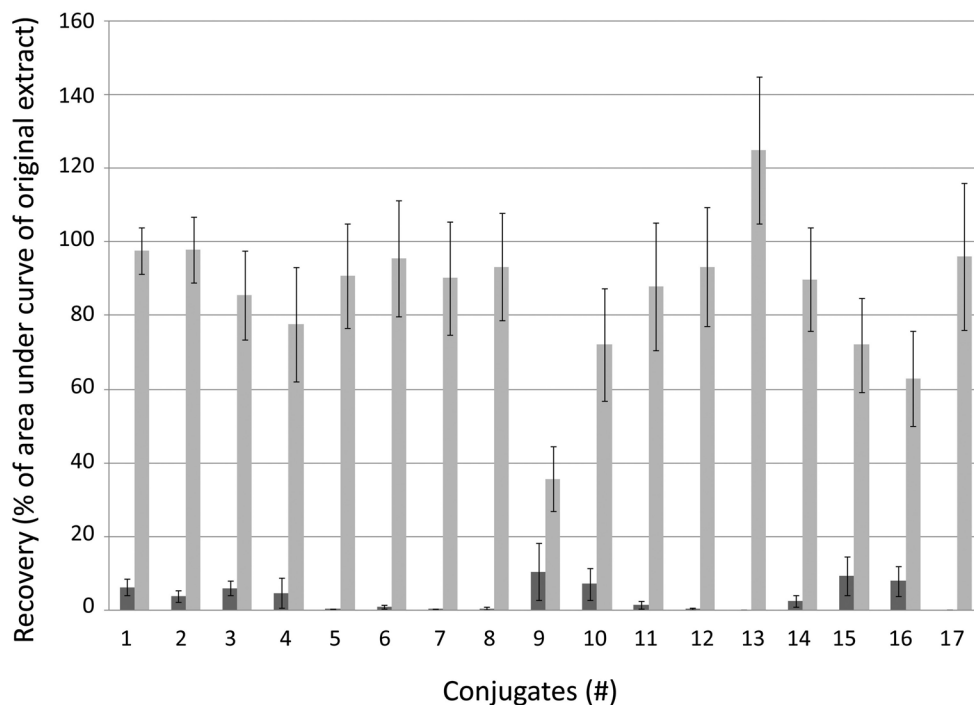
*$HPLC\text{-TOFMS}\text{-SPE}\text{-NMR}$  sample preparation:* One hour prior to the first test HPLC run, the freeze dried samples of eight SPE procedures (thus 12 ml of urine) were combined by redissolving in 800 microliters of 25% MeOH in  $\text{H}_2\text{O}$  containing 0.1% formic acid and filtering through a 0.45  $\mu\text{m}$  filter before transferring in an HPLC vial. The total concentration factor was 15.

*$Offline\text{-FTMS}^n$  sample preparation:* Hundred  $\mu\text{l}$  of the collected SPE fractions were dried at room temperature in the speedvac and redissolved in 100  $\mu\text{l}$  50% MeOD in  $\text{H}_2\text{O}$  containing 0.1% formic acid (FA). 12  $\mu\text{l}$  of redissolved fractions was subsequently transferred to a 384 wells plate and 5  $\mu\text{l}$  isopropanol was added to ensure spray stability.

*$HPLC\text{-FTMS}^n$  conditions:* Adaptations in  $HPLC\text{-FTMS}^n$  protocols compared to the protocol described earlier.<sup>1</sup> The following adapted gradient of  $\text{H}_2\text{O}$  containing 0.1% FA (solvent A) and acetonitrile containing 0.1% FA (solvent B) was applied using an injection volume of 5 microliters: at 0 min 5% B, at 85 min 50% B, at 87 min 75% B, at 92 min 75% B, at 95 min 5% B, and a 10 min

conditioning step of 5% B. Green tea urine HPLC-FTMS<sup>n</sup> data was obtained in negative ionization mode with a m/z window of 90-1200 Da in MS1. For black tea urine samples, the mass window (90-1500 Da) and gradient (5% B to 45% B in 85 min) were slightly different.

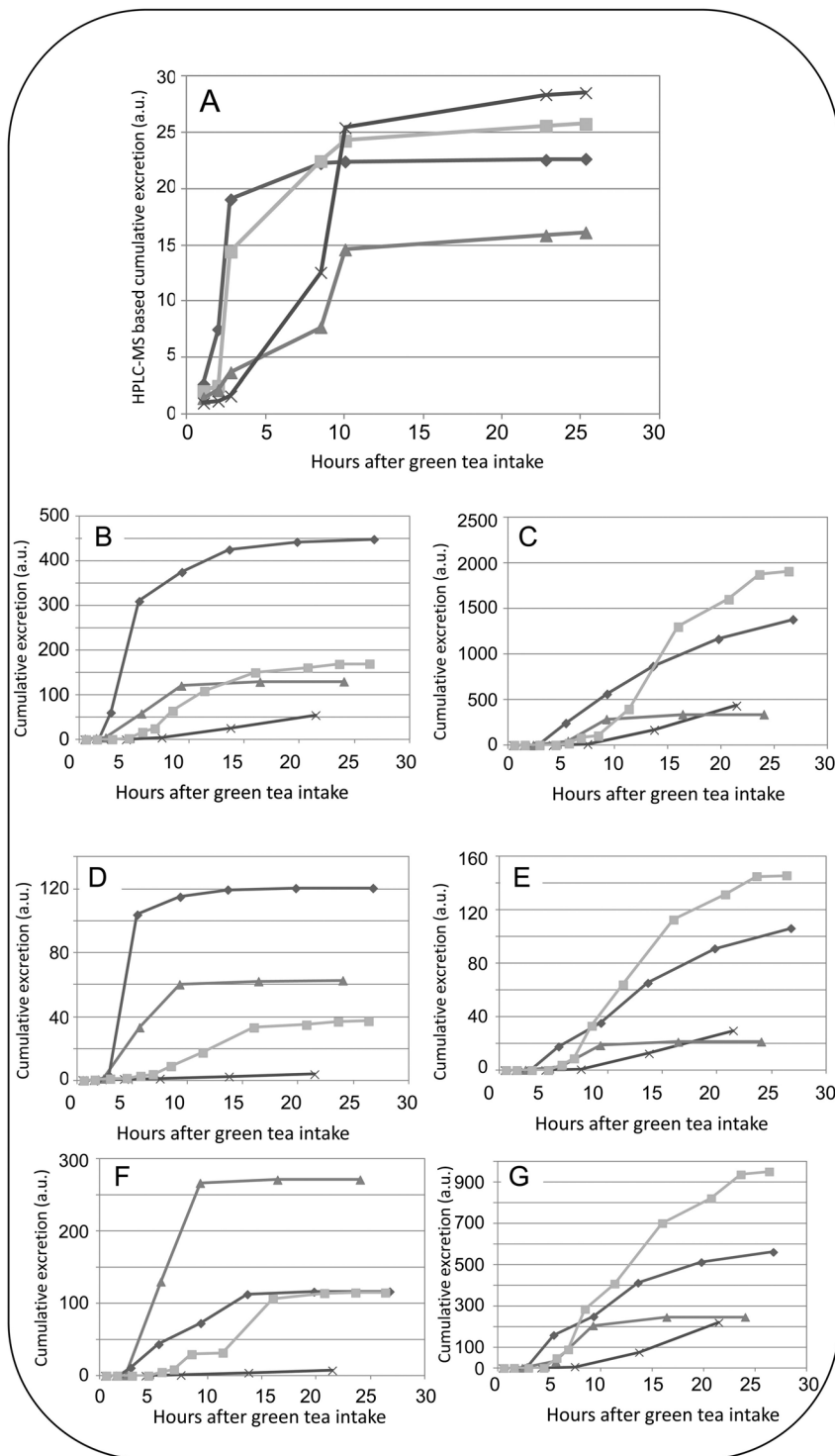
***Supporting Table 1:*** *This Excel file provides the spectral data of the annotated metabolites in this study. This Excel file is available at <http://edepot.wur.nl/216854> and a direct link is: <http://edepot.wur.nl/216861>*



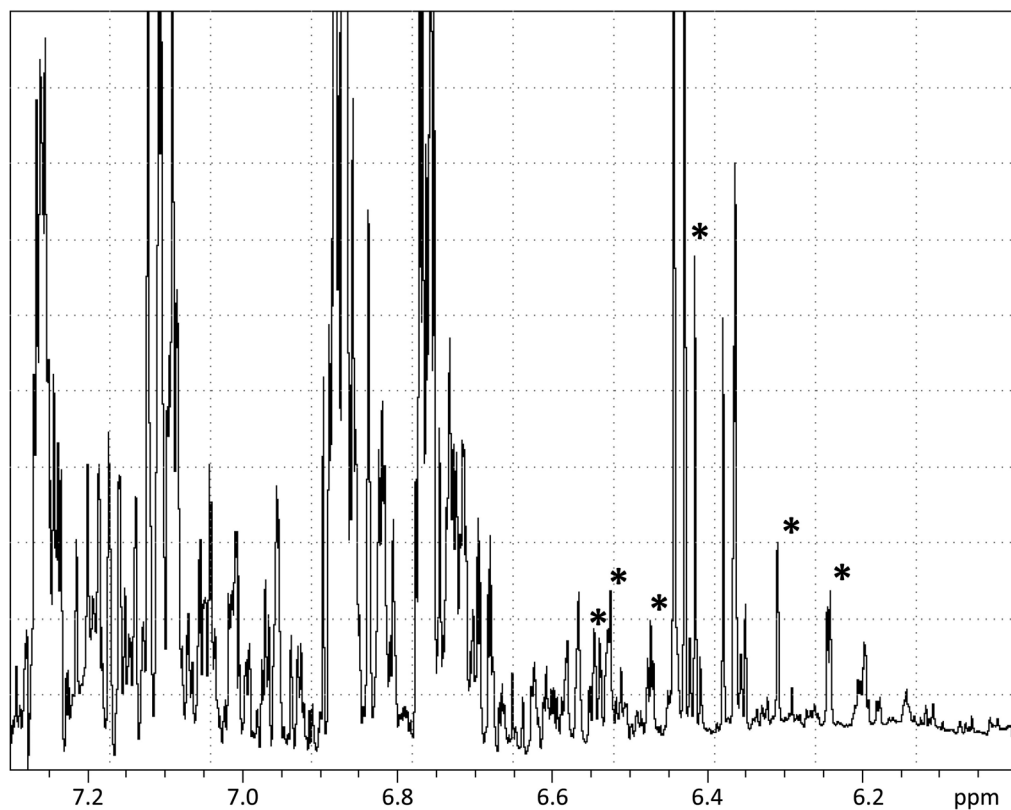
Conjugate (#)	Compound annotation	Mass (M-H)	RT (Q-ToF)
1	Hippuric acid	178	14.41
2	Phenylacetylglutamine (PAG)	263	15.26
3	p-Cresol-sulphate	187	26.7
4	Valerolactone2-glucuronide	383	10.62
5	Valerolactone2-glucuronide 2	383	15.28
6	Valerolactone2-glucuronide 3	383	16.72
7	Valerolactone-3-O-glucoronide	367	15.93
8	Valerolactone-4-O-glucuronide	367	18.6
9	Vanillic-acid-4-O-glucuronide	343	6.36

10	Vanillic-acid-COOH-O-glucuronide	343	9.61
11	Valerolactone 2-sulphate 1	287	16.7
12	Valerolactone 2-sulphate 2	287	22.19
13	Valerolactone 3-sulphate 1*	303	27.99
14	Methylphenyl-4-O-glucuronide	283	21
15	Pyrogallol-2-O-sulphate	205	13.09
16	Pyrogallol-2-O-glucuronide	301	11.36
17	valerolactone3-glucuronide 2*	399	23.54

**Figure S1:** Results of triplicate experiment of recoveries of urinary conjugates of phenolic acids and valerolactones during HLB 3cc SPE procedure. The HPLC-PDA-QTOFMS system as described by Moco et al.<sup>2</sup> (10  $\mu$ l injection volume) was used to determine the recoveries by the ratio's of peak heights in the wash and methanol elution fractions compared to peak heights in the initial urine extract. The recoveries of 17 conjugates in the wash (using water, ■) and in the methanol elution step (▣) are plotted. Error bars represent two times the standard deviation up and down of triplicate measurements. The two metabolites indicated with an asterisk (\*) were low abundant compounds and their error was set to 20%.

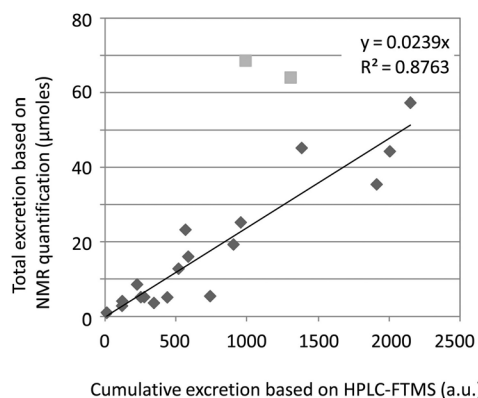


**Figure S2 (on left page):** HPLC-FTMS based cumulative urinary excretion of conjugates are presented. Relative quantification was performed based on peak areas in full scan HPLC-FTMS profiles. In plot A, four urinary conjugates are plotted that were excreted by volunteer 2 sampled over 26 hours, i.e. methylgallic acid I [◆], (Epi)catechin-O-sulphate-O-methyl III [■], 5-(4'-hydroxyphenyl)- $\gamma$ -valerolactone-4'-O-glucuronide [▲], and 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-O-glucuronide-O-sulphate [×]; Plots B-G show the HPLC-FTMS based cumulative excretion of 6 selected urinary conjugates in 4 volunteers (volunteer 1 [×], 2 [▲], 3 [◆], and 4 [■]), that is 5-(3',4'-dihydroxyphenyl)- $\gamma$ -valerolactone-4'-O-glucuronide (B), 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-sulphate (C), 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide (D), 4-hydroxy-5-(3',5'-dihydroxyphenyl)- $\gamma$ -valeric acid-3'-O-sulphate (E), 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-4'-O-glucuronide (F), 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide (G).



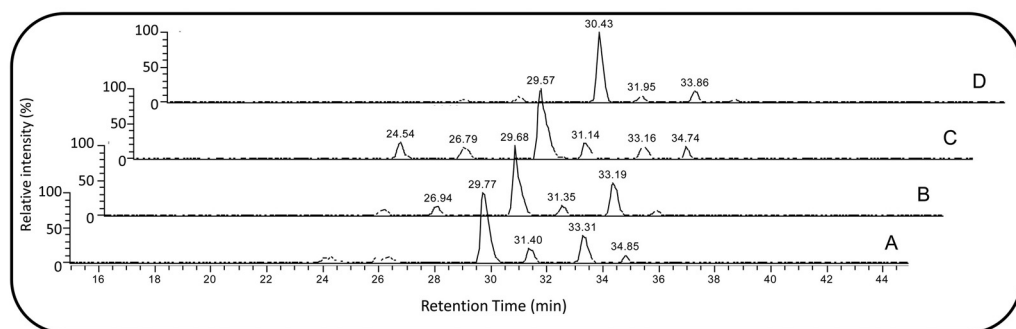
**Figure S3:** part of 1D-<sup>1</sup>H-NMR spectrum of HLB SPE prepurified and preconcentrated urine extract. Asterisks (\*) mark diagnostic NMR peaks used for quantification of valerolactone and pyrogallol conjugates, i.e., from lower to higher ppm values: (deconjugated) 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-sulphate [6.23 sd], 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide [6.31 s], pyrogallol-2-O-sulphate and pyrogallol-2-O-glucuronide [6.42/6.43 d], 5-(3',5'-dihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide [6.43 t], 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-glucuronide [6.52 bs], 5-(3',4',5'-trihydroxyphenyl)- $\gamma$ -valerolactone-3'-O-sulphate and its 4-hydroxyvaleric acid analogue [6.55/6.555 sd].





**Figure S4:** The correlation between the cumulative area under the curve of five conjugates for all time points in the four volunteers based on the relatively quantification with HPLC-FTMS<sup>n</sup> (x-axis) and the quantification based on 1D-<sup>1</sup>H-NMR (y-axis) is plotted. Linear regression was performed with and without two outlier points indicated with grey squares, resulting in R<sup>2</sup> values of 0.6386 and 0.8764 (trendline shown in plot) respectively. The trendline was fixed through the origin, which did not influence the R<sup>2</sup> significantly. The two outlier points were found to have significant overlapping peaks within the NMR peak area of the 1D-<sup>1</sup>H-NMR spectrum used for quantification.

The two outlier points were found to have significant overlapping peaks within the NMR peak area of the 1D-<sup>1</sup>H-NMR spectrum used for quantification.



**Figure S5:** Full scan HPLC-FTMS profiles of  $m/z$  383.0444 (3 ppm accuracy) in the retention window 15 – 45 min in the four volunteers (A-D). The  $m/z$  corresponds to the epicatechin-*O*-sulphate-*O*-methyl isomers.

### References in Supporting Information

1. J. J. J. Van der Hooft, J. Vervoort, R. J. Bino and R. C. H. de Vos, *Metabolomics*, Metabolomics 2012, **8** (4), 691-703.
2. S. Moco, R. J. Bino, O. Vorst, H. A. Verhoeven, J. de Groot, T. A. van Beek, J. Vervoort and C. H. R. de Vos, *Plant Physiol.*, 2006, **141**, 1205-1218.



## Chapter 7: General Discussion

### Structural elucidation of metabolites at the micromolar level

Current status in data acquisition and analysis  
for metabolite identification purposes

#### *Abstract*

Impressive new developments are ongoing in the structural elucidation of low abundant metabolites present in complex sample matrices. Recent literature and the work presented within this thesis show that hyphenation of mass spectrometry (MS) and one dimensional (1D)-proton nuclear magnetic resonance spectroscopy ( $^1\text{H-NMR}$ ) can provide evidence for complete structural elucidation of small molecules in complex matrices. These methods circumvent the need for extensive compound concentration as is required for two-dimensional (2D)-NMR experiments. These new methods facilitate accurate compound annotations. The further expansion of the structural elucidation of metabolites depends on the development of rich databases of both MS and NMR spectra, as well as the development of software that can perform candidate selection and rejection based on acquired spectral data.

#### *The role of compound identification in metabolomics studies.*

Technological advancements in mass spectrometry (MS) hardware based on high mass-resolution and increased sensitivity have boosted metabolite detection. The increased sensitivity and selectivity of mass detectors helps to generate comprehensive metabolite profiles of many diverse biological systems. This opens ways to annotate specific low abundant metabolites present in organs, tissues, or even in single cells.<sup>1</sup> To standardize the many existing protocols, the Metabolomics Standard Initiative (MSI) has defined protocols and standards

for metabolomics research including metabolite annotation and identification. Different levels of metabolite identification (MI) were proposed,<sup>2</sup> ranging from level 1 identifications, that is completely identified structures using at least 2 independent analytical methods, down to level 4 identification, that is the presence of a metabolite feature in a sample. Since its introduction, the number of publications reporting the MSI metabolite identification level of the annotated metabolites is steadily increasing. A full identification comprises of precise information on how all atoms in molecule are connected. However, the stereochemistry of metabolites is often not addressed in identification studies, as is illustrated by the commonly accepted 2D-representation of molecules. This is, at least partly, caused by the lack of available analytical tools and data experts to fully describe a molecule's three dimensional (3D)-structure. A complete identification of metabolites, including substitution positions and stereochemistry, is often needed to fully understand the role of small molecules in biological systems. For instance, recent studies on bioavailability and absorption of flavan-3-ol and flavanone isomers show that positional isomers and stereoisomers have different functional properties in humans.<sup>3-5</sup> Thus, robust and efficient metabolite identification strategies are needed to correctly elucidate and annotate the increasing amount of molecules detectable in metabolomics studies.

Metabolomics tools have become indispensable in many different scientific disciplines since their introduction a decade ago.<sup>6-11</sup> Metabolomics aims to qualitatively and quantitatively describe all metabolites in a given biological system.<sup>1, 6</sup> Large-scale metabolite identification is yet recognized as one of the major focus points in the metabolomics field.<sup>1, 12-17</sup> This is caused by the enormous chemical diversity in compounds, illustrated by the theoretical amount of  $6.2 \cdot 10^8$  elemental formulas (EFs) for small molecules (molecular weight <2000 Da) with the most common elements occurring in biological systems.<sup>18</sup> A single EF can represent numerous positional and stereochemical isomeric species, thereby expanding the number of theoretical existing metabolites even further to tens of billions. It has been estimated that 10% of the signals detected by current mass

spectrometry (MS) based metabolomics approaches, in both human and plant sciences, can be annotated to known entities<sup>19</sup> and that web-based compound libraries currently contain about 3% of the metabolites present in nature. Thus, there is definitively a need for systematic and (semi-)automatic approaches to achieve more level 1 MSI metabolite identifications in metabolomics studies.

In this paper we discuss the developments in metabolite identification and address the challenges and opportunities for the nearby future. We focus on spectral data acquisition and analysis with liquid chromatography (LC) coupled to both MS and nuclear magnetic resonance spectroscopy (NMR). This combination of technologies proves to be very effective for the identification of polar to semi-polar, small (that is <1500 Da) molecules containing the most common elements carbon (C), hydrogen (H), oxygen (O), nitrogen (N) and sulphur (S). Using the economically important class of (poly)phenolic compounds and their derivatives as an example, we review the currently used hyphenated platforms that combine MS and NMR measurements in terms of their sensitivity and practical use, followed by evaluating spectral data analysis approaches for structural elucidation including semi-automated compound annotation strategies.

*Acquisition of good quality metabolomics data for metabolite identification.*

*Sample preparation:* A sample preparation that is easy, fast, and robust is a prerequisite in acquiring metabolomics data sets. With the increased sensitivity of analytical detectors, reduction of contaminant signals and (electronic) background becomes important. The purity of compounds can be essential for complete structural elucidation, as abundant signals in MS or NMR from other compounds can suppress or mask the signals of the compound of interest, thereby hampering its correct assignment. In sample preparation steps, solid phase extraction (SPE) was introduced in the early 1970s and since then, new sorbents and applications have been constantly developed.<sup>20</sup> SPE combines sample (pre-) concentration and initial purification, as was recently shown in studies with plant and urine extracts (chapter 6 of this thesis).<sup>21</sup>

*Separation of metabolites:* Biological extracts usually contain thousands of

metabolites, including several isobaric compounds with the same EF. Therefore, adequate separation of compounds present in crude extracts is crucial to allow for the identification of closely related and closely eluting compounds. Prior to detection of compounds, the separation as provided by either liquid chromatography (LC) or gas chromatography (GC) will reduce the complexity of the background matrix thereby preventing overlapping MS and NMR signals as far as possible.<sup>12,22</sup>

*Detection of structural features of metabolites:* A Diode Array Detector (DAD) detects metabolites by their Ultra violet (UV) absorption and is normally placed directly after the chromatographic column (i.e. in-line) and provides useful structural information on chromophore-containing molecules that can be very helpful in the metabolite identification process.<sup>23</sup> However, when used after LC in complex extracts, UV lacks specificity and resolution as compared to MS. Often, biological extracts are so full of chromophore containing molecules, like phenolic acids and polyphenols, that metabolites are hard to differentiate solely on the basis of their UV characteristics. Hence, some chromophore groups with specific UV patterns, like urolithins, can be discriminated within complex extracts.<sup>24</sup> Therefore, LC-MS has become a major approach for profiling biological extracts containing non-volatiles like phenolic conjugates, flavonoid glycosides, and derivatives present in human urine. In current metabolomics studies, the mostly used detection platforms are MS<sup>25</sup> or NMR based.<sup>26,27</sup>

LC-MS based platforms generate thousands of so-called mass features comprising of parent ions, in-source fragments, and adducts, representing several hundreds of compounds.<sup>1,15,28</sup> The use of accurate mass based platforms facilitates fast EF assignment to detected metabolite features.<sup>29</sup> Knowing the EF is one of the key steps toward identification of a detected metabolite.<sup>18</sup> On top of that, the current performance of MS equipment in terms of mass resolving power minimizes mass overlap (avoiding confounding signal overlap), as compared to, for instance, LC-UV.<sup>1</sup>

*Integration of ultra pressure separation:* A recent development in the separation sciences was the step from HPLC toward Ultra Pressure/Performance Liquid

Chromatography (UPLC).<sup>30</sup> UPLC offers several advantages over HPLC, including an improved resolving power and a decreased analysis time. However, as a consequence of the frequently used shorter run times with a steeper solvent gradient, some isomers may separate badly or not at all using UPLC. Also, due to the relative high flow rate, UPLC usually generates very sharp chromatographic peaks with a lower amount of scans as compared to the slower HPLC separation. The integration of UPLC with accurate mass spectrometers, such as the Orbitrap FTMS is a focus point for the nearby future, especially when using MS<sup>n</sup> fragmentation on the fly.<sup>1,30,31</sup> Nevertheless, for high-throughput full scan metabolite profiling, UPLC offers sufficient advantages to become the preferred method over HPLC.<sup>30</sup> Recently, UPLC was combined with NMR in an indirect (at-line) manner. UPLC-TOFMS was used to isolate 5 – 10 µg of closely related compounds and spectral data obtained with capillary NMR was used to structurally elucidate wound stress biomarkers and several jasmonate related metabolites in *Arabidopsis thaliana*.<sup>32, 33</sup> The next opportunity is to further increase the mass sensitivity and to reduce the use of eluent solvents with the use of nano-LC technologies in metabolomics studies.<sup>34</sup>

*Fragmentation of metabolites in the mass spectrometer:* To increase the structural information obtained within a single LC-MS run, fragmentation of metabolites in the mass spectrometer can be used to detect compound class-specific fragment ions or fragment losses. In view of this, a more routine use of LC-MS/MS approaches has been advocated.<sup>15, 28</sup> However, care should be taken when directly comparing fragmentation spectra of different metabolites, especially when generated at different LC-based platforms, as unlike the high energy fragmentations used in GC-MS/MS, the low energy collision induced dissociation (CID) fragmentation, which is compatible with LC, may result in variable fragmentation spectra. Therefore, we developed a robust and accurate mass multistage MS (MS<sup>n</sup>) and LC-MS<sup>n</sup> approach, using a LTQ Ion trap - Orbitrap combination that generates so-called spectral trees.<sup>35, 36</sup> In these studies, detailed accurate mass MS<sup>n</sup> spectral trees were produced, thereby systematically fragmenting a metabolite into smaller fragment ions, resulting in fragment ions

and losses that can be specific for a compound class. Moreover, on the basis of parent-daughter ion relationships, different fragmentation pathways are usually revealed for the fragmented metabolite. Direct infusion of compounds to be identified was performed with the NanoMate robot. The use of the NanoMate robot has previously been reported as a promising tool in metabolomics studies as it needs only a few  $\mu\text{L}$  to produce a nano-electrospray into the MS for 30 minutes or more.<sup>15</sup>

The combination of technologies makes it possible to generate detailed  $\text{MS}^n$  spectral trees up to the  $\text{MS}^5$  level using 15 min of spraying (chapters 2 and 3 of this thesis).<sup>35,36</sup>  $\text{MS}^n$  data can also be obtained online, i.e., on-the-fly during the LC run, but due to the fact that recording online  $\text{MS}^n$  data is restricted to the chromatographic peak width of the eluting compound, these online spectral trees can be obtained in a robust manner up to  $\text{MS}^3$  level at most. Nevertheless, LC- $\text{MS}^n$  still made it possible to differentiate and identify several isomeric compounds by comparing the observed fragment peak ratios of metabolites within the extract and with the in-depth offline data, up to  $\text{MS}^5$ , previously generated from reference compounds.<sup>35</sup> For example, the polyphenolic cores of complex glycosides, and in some cases even the type of linkage of the sugar moieties, could be determined.<sup>36</sup>

The generated LC- $\text{MS}^n$  profiles proved to be rich sources of structural information from detected metabolites, since more than 400 metabolites, including nearly 70 novel metabolites, were annotated from LC- $\text{MS}^n$  runs of crude plant as well as human urine extracts (chapters 3, 5, and 6 of this thesis);<sup>21, 36</sup> however, for complete structure elucidation, without the availability of reference compounds, complementary information from NMR spectral data remains crucial.<sup>17</sup>

The reproducibility of the  $\text{MS}^n$  fragmentation approach between technical replicates, over time, at small variations in settings, and at different compound concentrations was found to be good, i.e., the standard deviation of the relative intensities of fragment ions was usually within 2%.<sup>35</sup> These results are an indication of the robustness of  $\text{MS}^n$  spectral tree data. On the basis of our results and the comparison with previous results from other groups generating nominal



mass MS<sup>n</sup> data of metabolites on Ion trap-MS machines,<sup>37, 38</sup> MS<sup>n</sup> mass spectral databases are a powerful annotation tool. The next step is to prove that MS<sup>n</sup> spectral trees are comparable between different Ion trap (hybrid) MS systems, and as a first step to test MS<sup>n</sup> fragmentation between different instruments, a reproducibility test between 4 different Ion trap - Orbitrap FTMS instruments (Round Robin) is ongoing within the Netherlands Metabolomics Centre.

*The use of hyphenated platforms to couple LC and NMR:* Several hyphenated systems combining LC with NMR have been developed to generate structural information to elucidate metabolites from crude extracts from both human and plant origin.<sup>17</sup> Over the last two decades, the fields of natural product research, drug discovery, and biomarker research boosted different directions of technological advancements in hyphenation and metabolomics analysis. In natural product research studies, the use of extensive 1D and 2D-NMR advanced to completely elucidate the very diverse compounds present in the biological extracts. One major drawback is that such studies usually require an extensive purification and concentration from at least a few tens of g dry weight of, e.g., plant sample to obtain sufficient input material. As an alternative, repeated injections of a plant extract on an analytical LC with UV-guided SPE-trapping of chromatographic peaks of interest, followed by elution of the trapped compound from the SPE cartridge into small probes for NMR was developed.<sup>39-41</sup> In drug discovery research, usually a single known compound is consumed or injected at high dose, followed by detection and identification of its metabolized products in body fluids. Technological advancements in terms of sensitivity of analytical detectors facilitate in the increasing demand to detect and identify all metabolized drug compounds. Here, as an example, hyphenated techniques have been applied to completely assess sulphated conjugates in urine.<sup>42</sup> In biomarker research in relation to, for instance, nutrition and diseases, two areas were mainly boosted. On the one hand, sophisticated statistical approaches have been developed to extract significant markers or features from large sets of MS or NMR spectral data.<sup>43</sup> On the other hand, targeted approaches have been optimized that allow for direct identification and quantification of statistically different markers.

The hyphenation of LC and NMR is not straightforward due to the non-deuterated solvents used in regular HPLC.<sup>40, 44</sup> The introduction of SPE-NMR did facilitate the connection between HPLC and NMR by trapping compounds of interest based on their UV characteristics on small SPE cartridges, thereby enabling the removal of non-deuterated solvents and elution with deuterated solvents that are suitable for NMR.<sup>40</sup> At the same time, analytes can be concentrated by performing multiple trapping runs, which is useful to collect sufficient analyte for NMR measurements or to decrease the measurement time needed. Structural information obtained from separately run LC-MS and UV guided SPE-NMR hyphenated platforms has been used for complete metabolite identification, as, for instance, shown by the identification of 22 secondary metabolites from *Kanahia laniflora* and *Harpagophytum procumbens* (Devil's claw)<sup>45, 46</sup> and the elucidation of two therapeutic agents using 15 µgram of analyte isolated from *in vitro* production with liver microsomes.<sup>47</sup>

The SPE trapping of analytes occurs upon triggering by a signal, for instance the UV absorption at 256 nm. Recently, instead of the use of the DAD signal as a trigger to trap, the MS signal was used.<sup>48</sup> The MS based SPE-NMR approach is an advantageous hyphenation for fast metabolite identification as it creates a direct link between MS and NMR data. Also, it allows more selective trapping and the trapping of non-chromophores as compared to UV based SPE-NMR (chapter 4 of this thesis).<sup>48</sup> Using the MS guided SPE trapping of analytes, the MS data enables candidate selection and rejection based on mass spectral characteristics and spectral data from the obtained 1D-<sup>1</sup>H-NMR spectrum of a trapped LC-MS peak is then often sufficient to discriminate between isomers based on their typical proton NMR patterns. For instance, to assess NMR proton patterns that are characteristic for ortho, para, or meta substitutions on a phenolic ring, only 1D-<sup>1</sup>H-NMR is sufficient.<sup>49</sup> To further illustrate the power of MS guided trapping, it was shown that with the help of only one 1D-<sup>1</sup>H-NMR literature reference spectrum, 3 structures, present in the NMR probe in amounts ranging from 200 ng to 2 µg, of tetraglycosylated and acylated quercetin and kaempferol conjugates were completely solved based on their LC-MS<sup>n</sup> and 1D-<sup>1</sup>H-NMR spectral data

sets (chapter 5 of this thesis).<sup>21</sup> As shown, this strategy is functioning well for phenolics, due to the availability of good quality spectral databases for this class of compounds and some of their unique properties, like displaying proton signals in the aromatic part of the NMR spectrum. The strategy can very well be extended to other classes of compounds by obtaining data sets of reference compounds and identifying diagnostic MS<sup>n</sup> and NMR features.

*From milligram to (sub)microgram level needed for compound identification.*

In general, with increasing m/z, the complexity of metabolites intensifies, for example because more combinations of elements are theoretically possible.<sup>18</sup> Thus, more structural information is needed for complete identification of larger metabolites. In practice, with 30 µg (300 µM in a 3 mm NMR tube [200 µL content] for MW of 500) of (>80%) pure compound in the probe, one can perform a complete set of NMR experiments that allows for a full structural elucidation of the compound. At lower amounts, i.e. 10 µg (100 µM), 1H-1H 2D experiments are still possible.<sup>17, 50</sup> As a result, crude plant extracts with rather concentrated metabolites can be directly injected into hyphenated systems to obtain (partially) purified NMR spectra of, for example, plant flavonoids and their glycosides.<sup>48</sup> Extracts containing less concentrated metabolites, i.e., low and submicromolar, that need to be elucidated will surely benefit from pre-concentration before HPLC trapping approaches, since otherwise the injection volumes would become larger than practically usable. For example, to obtain 10 µg in the NMR probe of a compound present at 1 µg/ml (that is 2 µM for a compound of 500 Da), one needs to inject 10 ml of extract. Upon sample concentration with a factor 15, for instance by using SPE, the injection volume needed reduces to 200 microliters, which can be achieved in 4 or 5 consecutive injections on an analytical LC system.

Combining the structural information obtained from LC-MS<sup>n</sup> and HPLC-MS-SPE-NMR results helps to identify low abundant compounds from a few hundred milligrams of crude sample. The combination of these platforms overcomes the use of 2D-NMR. Currently, microgram amounts of (partial)

pure trapped analytes are sufficient for 1D-<sup>1</sup>H-NMR. The use of SPE-trapped analytes in NMR analysis poses some challenges during data analysis, caused by dominant NMR signals of residual solvents and column material that may mask NMR peaks of the target analyte and that can hamper the identification of some metabolites. An advantage in the NMR analysis of phenolic molecules after SPE trapping is that they display characteristic NMR proton signals in the aromatic part of the NMR spectrum (i.e. 6 – 8 ppm), which is relatively clean as compared to the aliphatic part (1 – 2.5 ppm). For example, structural elucidation of aliphatic rings or lipids can be hindered by the presence of residual column materials, typically appearing as broad peaks in the 1 – 2 ppm part of the NMR spectrum. Phenolic glycosides display sugar protons in the 3.0 – 3.5 ppm area where residual protonated MeOD is also dominantly present in the spectrum. In practice, it is impossible to get rid of all the non-deuterated water in a sample; and therefore, the water signal is usually suppressed with special NMR pulse sequences. With sufficient amounts of analyte (i.e., >10 µgram), however, the structural information from 2D-<sup>1</sup>H-NMR analyses and LC-MS<sup>n</sup>, usually leads to correct peak assignments.

With less material and less analyte available, the implementation of 1.7 mm cryprobes for NMR analysis, as recently launched by Bruker, is promising, as this probe is a factor 4 to 5 more sensitive than the currently used 5 mm cryprobes. With these new probes, only 30 µliters of sample volume is needed, as compared to the 200 µliters sample volume needed for 3 mm NMR tubes. A recent application of the 1.7 mm cryprobe in combination with HPLC-SPE-NMR shows that within 7 hours measuring time, good quality <sup>13</sup>C-NMR spectra can be obtained from only 240 µgram (test extract) and 560 µgram (crude extract) of sample.<sup>51</sup> The <sup>13</sup>C-NMR spectra were used for the complete structural elucidation of several triterpenoids.<sup>51</sup> These results indicate that a further decrease of sample volume is possible. It is foreseen that the use of 1.7 mm cryprobes enables routine 1D-<sup>1</sup>H-NMR measurements of 100 nanogram amounts (200 picomol for a compound with m/z 500) within 30 minutes.

*Quantification of identified compounds.*

Quantification of metabolites was recently reviewed for different analytical platforms, i.e., GC-MS and LC-MS<sup>16, 52, 53</sup> and for NMR.<sup>14, 54</sup> According to these reports, LC-MS reference compounds are needed for quantification, especially for conjugated molecules, instead of the currently accepted method to take the parent compound response as representative for its glycosylated or metabolized products.<sup>55, 56</sup> A lack of reference compounds for secondary metabolites of plant origin and conjugated endogenous and exogenous metabolites in, for instance, urine extracts prevents fast characterization of metabolite fingerprints and accurate metabolite quantification. Trapping or fractionating chromatographic peaks followed by NMR-based metabolite identification and quantification gives the opportunity to recover the isolated analytes from the NMR probe and use the (partial) purified metabolites as reference compounds in LC-MS based experiments. This is an alternative approach to other existing strategies aiming for secondary metabolite production that use chemical synthesis or can produce plant glycosides in bacterial systems (Beekwilder et al., submitted), or human metabolite standards in microsomes.<sup>57</sup> Studies using the alternative approach of NMR based quantification of (partially) purified metabolites show promising results as they indicate that isolated drug metabolites can be used for MS quantification purposes.<sup>58-60</sup>

*Metabolite identification and annotation strategies.*

Two different strategies for the identification of a detected compound can be recognized: i) processing of the spectral data and extraction of relevant features, and then comparison of these features to database spectra, or ii) using known spectral features, i.e., characteristic signals or spectra of certain molecules or compound classes, as input to search for related compounds in raw or processed data files.

Extracted specific spectral features can be indicative for structural elements. For example, CO<sub>2</sub> losses observed in MS are correlated to the presence of free carboxyl groups in the fragmented molecules. Similarly, exact mass losses

indicative for hexoses, deoxyhexoses, sulphates, and glucuronides can be observed. The biological origin of the extract can thereby be used to search for specific conjugations, as, for instance, glucuronide and sulphate conjugations can be expected to be present in plasma and urine extracts. Also, approaches to determine the metabolite-likeness of candidate structures are in development to further aid in the metabolite identification process by narrowing down the window of possible candidates.<sup>61</sup>

The two identification strategies require 1) robust spectral data, 2) reproducible processing of all data, 3) considerable manual inputs of data experts, and 4) high quality spectral databases. Data processing of NMR data sets is feasible within vendor specific software sets like TopSpin (Bruker), while for LC-MS data sets, free software programs like Metalign and XCMS are now available for the scientific community.<sup>62, 63</sup> These software tools allow for (chromatographic) alignment of large sample sets and produce output that is compatible with statistical analysis software that can extract relevant features.<sup>43</sup> More specialized tools are also emerging, like the Multi-stage Elemental Formula (MEF) tool to assign elemental formulas to  $MS^n$  spectral trees,<sup>64</sup> which can be regarded as a first step in the processing of accurate mass  $MS^n$  data.

The number and content of MS and NMR spectral libraries is constantly growing, which facilitates compound database searching. For example, automated MS/MS and  $MS^n$  data processing and subsequent library comparison are under development at several places.<sup>65, 66</sup> The approach to annotate and identify metabolites through hierarchical  $MS^n$  spectral trees is valuable, provided that more comprehensive  $MS^n$  databases will become available. Undoubtedly, MS-based annotation will benefit from automated work flows that are currently under development, such as those within the Netherlands Metabolomics Centre and the Netherlands eScience Centre. These approaches will be highly appropriate for a high-throughput metabolite annotation and identification strategy that will greatly enhance the significance of conventional LC-MS metabolite profiling results since many more metabolites will be reliably annotated, thus allowing for greater biological insights.

With the development of NMR spectral databases and processing tools, automated structure annotation procedures based on NMR spectra are getting more within hand. For example, ACD NMR software, as used in the Scifinder compound database (<https://scifinder.cas.org/scifinder>), can predict NMR shifts of compounds that have not yet been experimentally obtained.<sup>67</sup> However, the ACD NMR software has its limitations in terms of precisely predicting a 1D-<sup>1</sup>H-NMR spectrum including shifts and coupling constants. Currently, within the Netherlands Metabolomics Centre, a NMR database for phenolic compounds is under development based on the PERCH NMR software (Mihaleva et al., submitted).<sup>68</sup> This software package uses a total line-shape fitting procedure, enabling accurate prediction and fitting of NMR spectra that allow for candidate selection based on 1D-<sup>1</sup>H-NMR spectra and verification of proton signal assignments.<sup>48, 69</sup> Also, the PERCH NMR software can better correct for small changes in NMR shifts that occur as a result of slight differences in temperature, pH, or other variables during acquisition of large sample series, as compared to the ACD software. The PERCH NMR software recently also showed promising applications for automation of the quantification of a selected number of metabolites in larger series of samples.<sup>70</sup> The resulting database will contain assigned NMR spectra, both experimental and predicted, of more than 6000 flavonoid molecules in different deuterated solvents (Mihaleva et al., submitted). One of the key features of this database is the consistent labelling of protons, thereby greatly facilitating fast comparison of NMR spectra. In the nearby future, automated annotation of NMR spectra can be extended to other classes of compounds and metabolized derivatives thereof. To illustrate this, an elegant *in silico* NMR-based approach to dereplicate and quantify structurally related triterpene cores based on their different methylation patterns has recently been reported, thereby using methyl NMR shifts found in literature as input for classification binary trees that allow for automated assignment of triterpene skeletons.<sup>71</sup>



*Towards automated annotation of de novo identified compounds.*

The metabolite identification process of metabolites of a well-defined compound class of which the structure is almost solved, often referred to as ‘known unknowns’, is merely based on pattern recognition, because core structures in the molecule give rise to specific and often unique MS and/or NMR signals. This is beneficial for computer aid tools that can search and compare patterns in the spectral data. However, in case of *de novo* identifications, the unexpected patterns observed in MS and/or NMR cause the time-consuming structural analysis. The step toward automated aid in *de novo* structural elucidation using MS and NMR spectral data still needs to be made. This is likely to take considerable time, as small metabolites pose challenges to prediction software due to their highly variable chemical structures. Small changes in a core molecule can result in large differences in retention times, MS<sup>n</sup> fragmentation patterns, and NMR chemical shifts. For example, upon MS fragmentation of flavonoids, rearrangements can take place that are not yet well predictable. Nevertheless, robustly and reproducibly generated fragmentation data will enhance the power of prediction models since the algorithms can be fine-tuned on the basis of even small differences in spectral data between structurally related molecules. Currently, annotation software tools can already be used as aid in: i) candidate selection and rejection, ii) MS fragment and NMR proton assignments, iii) structure confirmation, and iv) automatic annotations for compound classes and groups like flavonoids, triterpenes, phenylvalerolactones, and phenylvaleric acids.

*MS and NMR spectral databases.*

Once identified, annotated metabolites should be stored in an efficient way that enables researches to quickly retrieve the relevant compound information they need. Several initiatives have led to the development of web-based platforms where metabolites can be queried based on their physic-chemical properties and spectral data from UV, MS, or NMR.<sup>72-75</sup> The number of compound databases is still expanding, for example with libraries dedicated to a specific organism<sup>76</sup> or specific parts of an organism like human serum.<sup>77</sup> However, fast and reliable



identification of already annotated metabolites still remains a considerable effort, for example due to differences in metabolite naming and different experimental conditions between databases.<sup>16, 71</sup>

In practice, spectral databases filled with MS and NMR data are most promising for metabolite identification purposes, as they allow for pre-selection of candidate structures based on mass values or elemental formulas, as well as fragmentation data, if present. Subsequently, further candidate selection can be performed based on NMR proton shifts, leading to the correct annotation of, for example, conjugation sites or the correct identity of an attached sugar moiety. Finally, the experimental NMR spectra can be fitted to the predicted NMR spectra of a few remaining candidate structures for the structural verification and as guide for systematic proton NMR assignments. However, in case of stereoisomers such as enantiomers, additional analytical information like circular dichroism is needed to completely characterize the stereochemistry of chiral metabolites.<sup>78, 79</sup>

### *Outlook*

The use of LC-MS<sup>n</sup> and HPLC-MS-SPE-NMR in combination with a SPE pre-concentration and pre-purification step is a step forward in metabolite identification of low abundant molecules in complex biological matrices and the approach is foreseen to be implemented in large-scale metabolomics studies applying accurate mass UPLC-MS<sup>2</sup> for initial screening. After selection of samples for more detailed HPLC-MS<sup>n</sup> measurements and extracting relevant mass features, masses of target compounds and/or major LC-MS base peaks that potentially contain characteristic metabolites in sufficient amounts can be selected in a few samples to obtain 1D-<sup>1</sup>H-NMR spectra using HPLC-MS-SPE-NMR. The available MS and NMR spectral databases will then drastically reduce the number of candidate structures for target molecules due to constraints by generated MS<sup>n</sup> fragments and conjugation patterns derived from NMR spectra. Furthermore, the approach reduces the need for elaborate 2D-NMR spectra, as similar structural information can be obtained from less amounts of analyte. One of the current limitations of the described strategy is that the obtained

structural information is not always sufficient to solve the chemical structure of yet unknown molecules or complex molecules, i.e., true *de novo* identification. Nevertheless, the increasing sensitivity of both MS and NMR facilitates *de novo* metabolite identification by signal gain, thereby providing more possibilities for detailed MS<sup>n</sup> and 2D-NMR experiments. At the same time, the sensitivity gain asks for dedicated sample preparation to reduce the background matrix as efficient as possible, because upon sample concentration, contaminants can also be present in elevated amounts. Therefore, we expect more dedicated sample preparation techniques to become available in the nearby future, for example using new sorbents for SPE, as the successful metabolite annotation and identification largely depends upon the purity of samples.

We also foresee major developments in the computational sciences that will support both LC-MS and NMR based metabolite identification and annotation toward computer assisted identification (C-AID) with the help of spectral references, like the above-mentioned proof of principle studies show. Currently, protocols are being established and defined for robust annotation of spectral data. It is expected that within several years, good quality spectral data, acquired on different analytical platforms, of representative molecules belonging to numerous compound classes will become available. All the recent developments in sample preparation, hyphenated data acquisition, and automated annotation should contribute toward faster data analyses in metabolite identification. However, the chemical creativity of Mother Nature is undoubtedly such large that for many not yet discovered metabolites still the input from the human brain will be needed for structural elucidation.

---

*References*

1. S. Kueger, D. Steinhauser, L. Willmitzer and P. Giavalisco, *Plant J.*, 2012, 70, 39-50.
2. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden and M. R. Viant, *Metabolomics*, 2007, 3, 211-221.
3. W. Brand, B. Oosterhuis, P. Krajcsi, D. Barron, F. Dionisi, P. J. Van Bladeren, I. M. C. M. Rietjens and G. Williamson, *Biopharmac. Drug Dispos.*, 2011, 32, 530-535.
4. W. Brand, J. Shao, E. F. Hoek-Van Den Hil, K. N. Van Elk, B. Spenkelink, L. H. J. De Haan, M. J. Rein, F. Dionisi, G. Williamson, P. J. Van Bladeren and I. M. C. M. Rietjens, *J. Agr. Food Chem.*, 2010, 58, 6119-6125.
5. J. I. Ottaviani, T. Y. Momma, C. Heiss, C. Kwik-Urbe, H. Schroeter and C. L. Keen, *Free Rad. Biol. Med.*, 2011, 50, 237-244.
6. R. Hall, M. Beale, O. Fiehn, N. Hardy, L. Sumner and R. Bino, *Plant Cell*, 2002, 14, 1437-1440.
7. J. K. Nicholson, J. Connelly, J. C. Lindon and E. Holmes, *Nat. Rev. Drug Discov.*, 2002, 1, 153-161.
8. A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart and S. Wopereis, *Metabolomics*, 2009, 1-24.
9. W. B. Dunn, *Phys. Biol.*, 2008, 5, art. no. 011001.
10. R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino and R. D. Hall, *Nat. Prot.*, 2007, 2, 778-791.
11. R. Mohamed, E. Varesio, G. Ivosev, L. Burton, R. Bonner and G. r. Hopfgartner, *Anal. Chem.*, 2009, 81, 7677-7694.
12. S. Moco, R. J. Bino, R. C. H. De Vos and J. Vervoort, *Tr. Anal. Chem.*, 2007, 26, 855-866.
13. F. Puiggròs, R. Solà, C. Bladé, M. J. Salvadó and L. Arola, *J. Chromatogr., A*, 2011, 1218, 7399-7414.
14. D. S. Wishart, *Bioanalysis*, 2011, 3, 1769-1782.
15. T. Kind and O. Fiehn, *Bioanal. Rev.*, 2010, 2, 23-60.
16. J. F. Xiao, B. Zhou and H. W. Ransom, *Tr. Anal. Chem.*, 2012, 32, 1-14.
17. J. L. Wolfender, G. Marti and E. F. Queiroz, *Curr. Org. Chem.*, 2010, 14, 1808-1832.
18. T. Kind and O. Fiehn, *Bmc Bioinf.*, 2007, 8, art. no. 105.
19. B. P. Bowen and T. R. Northen, *J. Am. Soc. Mass Spectrom.*, 2010, 21, 1471-1476.
20. L. Berrueta, B. Gallo and F. Vicente, *Chromatographia*, 1995, 40, 474-483.
21. J. J. J. van der Hoof, M. Akermi, F. Y. Ünü, V. Mihaleva, V. G. Roldan, R. J. Bino, R. C. H. De Vos and J. Vervoort, *J. Agr. Food Chem.*, 2012, **Article ASAP - DOI 10.1021/jf300297y**.

22. B. Álvarez-Sánchez, F. Priego-Capote and M. D. Luque de Castro, *Tr. Anal. Chem.*, 2010, 29, 111-119.
23. S. Moco, J. Vervoort, R. J. Bino, R. C. H. De Vos and R. Bino, *Tr. Anal. Chem.*, 2007, 26, 855-866.
24. R. González-Barrio, P. Truchado, H. Ito, J. C. Espín and F. A. Tomás-Barberán, *J. Agr. Food Chem.*, 2011, 59, 1152-1162.
25. K. Dettmer, P. A. Aronov and B. D. Hammock, *Mass Spectrom. Rev.*, 2007, 26, 51-78.
26. M. Coen, E. Holmes, J. C. Lindon and J. K. Nicholson, *Chem. Res. Toxicol.*, 2008, 21, 9-27.
27. H. K. Kim, Y. H. Choi and R. Verpoorte, *Tr. Biotechnol.*, 2011, 29, 267-275.
28. E. Werner, J. F. Heilier, C. Ducruix, E. Ezan, C. Junot and J. C. Tabet, *J. Chromatogr., B*, 2008, 871, 143-163.
29. T. Zhang, D. J. Creek, M. P. Barrett, G. Blackburn and D. G. Watson, *Anal. Chem.*, 2012, 84, 1994-2001.
30. R. Nicoli, S. Martel, S. Rudaz, J. L. Wolfender, J. L. Veuthey, P. A. Carrupt and D. Guillarme, *Expert Opinion Drug Discov.*, 2010, 5, 475-489.
31. A. Makarov and M. Scigelova, *J. Chromatogr., A*, 2010, 1217, 3938-3945.
32. G. Glauser, J. Boccard, S. Rudaz and J. L. Wolfender, *Phytochem. Anal.*, 2010, 21, 95-101.
33. G. Glauser, D. Guillarme, E. Grata, J. Boccard, A. Thiocone, P.-A. Carrupt, J.-L. Veuthey, S. Rudaz and J.-L. Wolfender, *J. Chromatogr., A*, 2008, 1180, 90-98.
34. J. Hernández-Borges, Z. Aturki, A. Rocco and S. Fanali, *J. Sep. Sci.*, 2007, 30, 1589-1610.
35. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino, J. Beekwilder and R. C. H. De Vos, *Anal. Chem.*, 2011, 83, 409-416.
36. J. J. J. Van der Hoof, J. Vervoort, R. J. Bino and R. C. H. de Vos, *Metabolomics*, 2012, 8, 691-703.
37. M. N. Clifford, K. L. Johnston, S. Knight and N. Kuhnert, *J. Agr. Food Chem.*, 2003, 51, 2900-2911.
38. F. Kuhn, M. Oehme, F. Romero, E. Abou-Mansour and R. Tabacchi, *Rapid Commun. Mass Spectrom.*, 2003, 17, 1941-1949.
39. V. Exarchou, M. Godejohann, T. A. v. Beek, I. P. Gerotheranassis and J. Vervoort, *Anal. Chem.*, 2003, 75, 6288-6294.
40. J. W. Jaroszewski, *Planta Med.*, 2005, 71, 795-802.
41. H. Tang, C. Xia and Y. Wang, *Magn. Res. Chem.*, 2009, 47, S157-S162.
42. J. P. Shockcor, S. E. Unger, I. D. Wilson, P. J. D. Foxall, J. K. Nicholson and J. C. Lindon, *Anal. Chem.*, 1996, 68, 4431-4435.
43. M. M. W. B. Hendriks, F. A. Eeuwijk, R. H. Jellema, J. A. Westerhuis, T. H. Reijmers, H. C. J. Hoefsloot and A. K. Smilde, *Tr. Anal. Chem.*, 2011, 30, 1685-1698.

44. J. W. Jaroszewski, *Planta Med.*, 2005, **71**, 691-700.
45. C. Clarkson, D. Staerk, S. H. Hansen, P. J. Smith and J. W. Jaroszewski, *J. Nat. Prod.*, 2006, **69**, 1280-1288.
46. C. Clarkson, D. Staerk, S. Honoré Hansen and J. W. Jaroszewski, *Anal. Chem.*, 2005, **77**, 3547-3553.
47. F. Gillotin, P. Chiap, M. Frédérick, J. C. Van Heugen, P. Francotte, P. Lebrun, B. Pirotte and P. De Tullio, *Drug Metab. Dispos.*, 2009, **38**, 232-240.
48. J. J. J. Van der Hooft, V. Mihaleva, R. J. Bino, R. C. H. de Vos and J. Vervoort, *Magn. Res. Chem.*, 2011, **49**, S55-S60.
49. J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, W. Wang, X. Lu, H. U. Häring, E. D. Schleicher, R. Lehmann and G. Xu, *Anal. Chem.*, 2008, **80**, 1280-1289.
50. T. F. Molinski, *Nat. Prod. Rep.*, 2010, **27**, 321-329.
51. S. G. Wubshet, K. T. Johansen, N. T. Nyberg and J. W. Jaroszewski, *J. Nat. Prod.*, 2012.
52. M. Koek, R. Jellema, J. van der Greef, A. Tas and T. Hankemeier, *Metabolomics*, 2011, **7**, 307-328.
53. M. R. Meyer and H. H. Maurer, *Anal. Bioanal. Chem.*, 2012, **402**, 195-208.
54. G. F. Pauli, T. Gödecke, B. U. Jaki and D. C. Lankin, *J. Nat. Prod.*, 2012, **75**, 834-851.
55. U. P. Dahal, J. P. Jones, J. A. Davis and D. A. Rock, *Drug Metabol. Dispos.*, 2011, **39**, 2355-2360.
56. V. Pieri, S. Sturm, C. Seger, C. Franz and H. Stuppner, *Metabolomics*, 2012, **8**, 335-346.
57. M. Clements and L. Li, *Anal. Chimica Acta*, 2011, **685**, 36-44.
58. A. Mutlib, R. Espina, K. Vishwanathan, K. Babalola, Z. Chen, C. Dehnhardt, A. Venkatesan, T. Mansour, I. Chaudhary, R. Talaat and J. A. Scatina, *Drug Metabol. Dispos.*, 2011, **39**, 106-116.
59. G. S. Walker, T. F. Ryder, R. Sharma, E. B. Smith and A. Freund, *Drug Metabol. Dispos.*, 2011, **39**, 433-440.
60. R. Espina, L. Yu, J. Wang, Z. Tong, S. Vashishtha, R. Talaat, J. Scatina and A. Mutlib, *Chem. Res. Toxicol.*, 2009, **22**, 299-310.
61. J. E. Peironcely, T. Reijmers, L. Coulier, A. Bender and T. Hankemeier, *PLoS ONE*, 2011, **6**, e28966.
62. A. Lommen, *Anal. Chem.*, 2009, **81**, 3079-3086.
63. R. Tautenhahn, G. J. Patti, D. Rinehart and G. E. Siuzdak, *Anal. Chem.*, 2012.
64. M. Rojas-Chertó, P. T. Kasper, E. L. Willighagen, R. J. Vreeken, T. Hankemeier and T. H. Reijmers, *Bioinformatics*, 2011, **27**, 2376-2383.
65. F. Rasche, K. Scheubert, F. Hufsky, T. Zichner, M. Kai, A. Svatoš and S. Böcker, *Anal. Chem.*, 2012, **84**, 3417-3426.
66. M. Rojas-Cherto, J. E. Peironcely, P. T. Kasper, J. J. J. van der Hooft, R. C. H. De Vos, R. J. Vreeken, T. Hankemeier and T. Reijmers, *Anal. Chem.*, 2012, **84**, 5524-5534.

67. M. E. Elyashberg, A. J. Williams and G. E. Martin, *Progr. Nucl. Magn. Res. Spectr.*, 2008, **53**, 1-104.
68. R. Laatikainen, M. Niemitz, U. Weber, J. Sundeun, T. Hassinen and J. Vepsäläinen, *J. Magn. Res. Ser. A*, 1996, **120**, 1-10.
69. M. A. Muñoz, N. Perez-Hernandez, M. W.ertino, G. Schmeda-Hirschmann and P. Joseph-Nathan, *J. Nat. Prod.*, 2012, **75**, 779-783.
70. J. G. Napolitano, T. Gödecke, M. F. Rodríguez-Brasco, B. U. Jaki, S. N. Chen, D. C. Lankin and G. F. Pauli, *J. Nat. Prod.*, 2012, **75**, 238-248.
71. F. Qiu, A. Imai, J. B. McAlpine, D. C. Lankin, I. Burton, T. Karakach, N. R. Farnsworth, S. N. Chen and G. F. Pauli, *J. Nat. Prod.*, 2012, **75**, 432-443.
72. T. Tohge and A. R. Fernie, *Phytochemistry*, 2009, **70**, 450-456.
73. A. Scalbert, C. Andres-Lacueva, M. Arita, P. Kroon, C. Manach, M. Urpi-Sarda and D. Wishart, *J. Agr. Food Chem.*, 2011, **59**, 4331-4348.
74. D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. de souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhtudinov, L. Li, H. J. Vogel and I. Forsythei, *Nucleic Acids Res.*, 2009, **37**.
75. M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-Mcintyre, P. Begley, K. Carroll, D. Broadhurst, A. Tseng, N. Swainston, I. Spasic, R. Goodacre and D. B. Kell, *Analyst*, 2009, **134**, 1322-1332.
76. T. Jewison, C. Knox, V. Neveu, Y. Djoumbou, A. C. Guo, J. Lee, P. Liu, R. Mandal, R. Krishnamurthy, I. Sinelnikov, M. Wilson and D. S. Wishart, *Nucleic Acids Res.*, 2012, **40**, D815-D820.
77. N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T. L. Pedersen, S. R. Smith, F. Bamforth, R. Greiner, B. McManus, J. W. Newman, T. Goodfriend and D. S. Wishart, *PLoS ONE*, 2011, **6**.
78. J. R. Kesting, L. Olsen, D. Staerk, M. V. Tejesvi, K. R. Kini, H. S. Prakash and J. W. Jaroszewski, *J. Nat. Prod.*, 2011, **74**, 2206-2215.
79. K. Sprogøe, D. Staerk, H. L. Ziegler, T. H. Jensen, S. B. Holm-Møller and J. W. Jaroszewski, *J. Nat. Prod.*, 2008, **71**, 516-519.







## *Chapter 8A: Summary*

Detailed knowledge of the chemical content of organisms, organs, tissues, and cells is needed to fully characterize complex biological systems. The high chemical variety of compounds present in biological systems is illustrated by the presence of a large variety of compounds, ranging from apolar lipids, semi-polar phenolic conjugates, toward polar sugars. A molecule's chemical structure forms the basis to understand its biological function. The chemical identification process of small molecules (i.e., metabolites) is still one of the major focus points in metabolomics research. Actually, no single analytical platform exists that can measure and identify all existing metabolites. In this thesis, two analytical techniques that are widely used within metabolite identification studies have been combined, i.e., mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR). MS was used to ionize the metabolites and to record their molecular weight and to provide substructure information based on fragmentation in the mass spectrometer. NMR gave the comprehensive structural information on the chemical environment of protons and their linkage to other protons within the molecule. The additional structural information as compared to MS is at the cost of an increased amount of compound needed for NMR detection and spectra generation. Here we combined both analytical methods into a liquid chromatography (LC)-based platform that concentrated compounds based on their specific mass; thereby providing a direct link between MS and NMR data. Another platform was developed that generated robust multistage MS<sup>n</sup> data, i.e., the systematic fragmentation of metabolites and subsequent fragmentation of resulting fragments.

This thesis aims to accelerate metabolite identification of low abundant plant and human derived compounds by following a systematic approach. The acquired structural information from MS<sup>n</sup> and 1D-<sup>1</sup>H-NMR spectra resulted in the complete elucidation of phenolic metabolites in microgram scale from both plant and human origin.

In the *chapter 1*, the analytical techniques and terms used throughout the thesis are introduced. *The second chapter* describes how a high mass resolution MS<sup>n</sup> fragmentation approach was tested in both negative and positive ionization modes for differentiation and identification of metabolites, using a series of 121 polyphenolic molecules. An injection robot was used to infuse the reference compounds one by one into a hybrid mass spectrometer, combining MS<sup>n</sup> possibilities with accurate mass read-out. This approach resulted in reproducible and robust MS<sup>n</sup> fragmentation trees up to MS<sup>5</sup>, which were differential even for closely related compounds. Accurate MS<sup>n</sup>-based spectral trees were shown to be robust and powerful to distinguish metabolites with similar elemental formula (i.e., isomers), thereby assisting compound identification and annotation in complex biological samples. In *the third chapter*, we tested the annotation power of this spectral tree approach for annotation of phenolic compounds in crude extracts from *Lycopersicum esculentum* (tomato) and the model plant *Arabidopsis thaliana*. Partial MS<sup>n</sup> spectral trees were generated directly after chromatographic elution (LC-MS<sup>n</sup>). Detailed MS<sup>n</sup> spectral trees could be recorded with the use of a collector/injector robot. We were able to discriminate flavonoid glycosides based on their unique MS<sup>n</sup> fragmentation patterns in either negative or positive ionization mode. Following this approach, we could annotate 127 metabolites in the tomato and *Arabidopsis* extracts, including 21 novel metabolites. The good quality MS<sup>n</sup> spectral trees obtained can be used to populate MS<sup>n</sup> databases and the protocols to generate the spectral trees are a good basis to further expand this database with more diverse compounds.

*Chapter 4* then describes how an automated platform, coupling chromatography with MS and NMR (LC-MS-solid phase extraction-NMR), was developed that can trap and transfer metabolites based on their mass values from a complex biological extract in order to obtain NMR spectra of the trapped LC-MS peak, out of minute amounts of sample and analyte. Extracts from tomatoes modified in their flavonoid biosynthesis pathway were used as proof of principle for the metabolite identification process. This approach resulted in the complete structural elucidation of 10 flavonoid glycosides. This study shows that improving the link

between the mass signals and NMR peaks derived from the selected LC-MS peaks decreases the time needed for elucidation of the metabolite structures. In addition, automated 1D-<sup>1</sup>H-NMR spectrum fitting of the experimental data obtained in this study using the PERCH NMR software further speeded up the candidate rejection process.

*Chapter 5* illustrates how the two developed analytical platforms could be used for the successful selection, annotation, and identification of 177 phenolic compounds present in different extracts of *Camellia sinensis*, i.e., green, white, and black tea extracts, including the full identification of microgram amounts of complex acylated conjugates of kaempferol and quercetin. Principal component analysis based on the relative abundance of the annotated phenolic compounds in 17 commercially available black, green and white tea products separated the black teas from the green and white teas, thereby illustrating the differential phenolic metabolite contents of black tea as compared to green and white teas. The change in phenolic profiles reflects the polymerization reactions occurring upon transformation of green tea into black tea. This study shows that the combined use of MS<sup>n</sup> spectral trees and LC-MS-solid phase extraction-NMR leads to a more comprehensive metabolite description thereby facilitating the comparison of tea and other plant samples.

*In chapter 6*, we aimed to structurally elucidate and quantify polyphenol-derived conjugates present in the human body by studying the urinary excretion of these conjugates. We applied a combination of a solid phase extraction preparation step and the two HPLC-coupled analytical platforms as described in chapters 2 and 3. This analytical strategy resulted in the annotation of 138 urinary metabolites including 36 completely identified valerolactone conjugates. These valerolactones are microbial breakdown products of tea phenols. NMR predictions of glucuronidated and sulphonated core metabolites were performed in order to confirm the NMR peak assignments on the basis of 1D-<sup>1</sup>H-NMR data only. In addition, 26 hours quantitative excretion profiles for certain valerolactone conjugates were obtained using diagnostic proton signals in the 1D-<sup>1</sup>H-NMR spectra of urine fractions.

In the *seventh chapter*, the current state of metabolite identification and expected challenges in the structural elucidation of metabolites at (sub)microgram amounts are discussed. The work in this thesis and of other groups working on the hyphenation of MS and NMR shows that the complete *de novo* identification of microgram amounts and even lower of compound is feasible by using MS guided solid phase extraction trapping in combination with 1D-<sup>1</sup>H-NMR or UPLC-TOF-MS isolation followed by capillary NMR. Semi-automated annotation of compounds based on their MS and NMR features is now feasible for some well studied compound classes and groups.

Altogether, the developed platforms yield new and improved insights in the phenolic profiles of well-studied plants as well as a comprehensive picture of the metabolic fate of green tea polyphenols upon intake in the human body. The followed metabolite identification strategy is also useful for other studies that aim to elucidate bioactive compounds, especially when only small sample volumes are available. In addition, this thesis contributes to the acquisition of good quality data for metabolite identification by acquiring robust MS<sup>n</sup> fragmentation spectra and 1D-<sup>1</sup>H-NMR spectra of partial purified analytes at microgram scale, which paves the path for further developments in data acquisition and analysis, as well as the unravelling of yet unknown metabolites in a faster, more systematic, and more automated manner.





## *Hoofdstuk 8B: Samenvatting*

Planten vormen een belangrijke bron van voedsel en zitten vol organische verbindingen met uiteenlopende functies in de plant zelf, maar ook, eenmaal opgenomen, in het menselijk lichaam. Biologisch actieve verbindingen hebben in de plant functies als bescherming tegen ultraviolet licht van de zon en pigmentatie, terwijl ze in de mens ontstekingsremmend kunnen zijn. Deze verbindingen, kleine moleculen, zijn opgebouwd uit een combinatie van enkele veel voorkomende elementen (atomen), zoals koolstof (C), waterstof (H), zuurstof (O) en stikstof (N) en andere minder vaak voorkomende elementen als fosfor (P) en zwavel (S). De opheldering van de chemische structuur van deze verbindingen is een lastige puzzel, omdat de atomen op vele manieren met elkaar verbonden kunnen zijn. Die verschillende mogelijkheden leiden dan ook tot een wijd scala aan verbindingen (de geschatte aantallen lopen uiteen van 200,000 tot 1,000,000) die aanwezig zijn in planten en mensen. De structuren van deze verbindingen lopen uiteen van kleine organische zuren en suikers, via iets grotere flavonoïden, alkaloiden, en carotenoïden, tot nog grotere vetachtige verbindingen (lipiden). Behalve in grootte, verschillen deze stoffen ook in andere eigenschappen. Zo lossen suikers makkelijk op in water (polair), terwijl lipiden juist moeilijk oplosbaar zijn in water (apolair). Het is dan ook niet verbazingwekkend dat er tot nu toe geen universele methode is die alle bestaande verbindingen tegelijkertijd kan analyseren. In recente studies die als doel hebben kleine moleculen te detecteren en hun chemische structuur op te helderen, worden voornamelijk twee analytische detectiemethoden gebruikt, vaak in combinatie met chromatografie en andere opzuiveringsmethoden: mass spectrometry (MS) en kernspin resonantie (NMR). MS detecteert geïoniseerde verbindingen en maakt het mogelijk om via de massa van de moleculen de atoomsamenstelling (elementformule) te bepalen. Hoe precieser de massa's gemeten worden, hoe betrouwbaarder de elementformule bepaald kan worden. Dit is een belangrijke eerste stap in molecuulidentificatie. Fragmentatie van moleculen in de MS levert informatie over substructuren van de gefragmenteerde verbindingen op.

Voor volledige opheldering van de chemische structuur is ook structuurinformatie van proton ( $^1\text{H}$ ) NMR nodig. NMR is minder gevoelig dan MS (meer stof is nodig voor detectie), maar geeft meer gedetailleerde informatie over hoe protonen gegroepeerd zijn binnen een molecuul. Door de structuurinformatie van MS en NMR te combineren moet het mogelijk zijn om de moleculaire puzzel op te lossen. De analysetechnieken voor de detectie en chemische opheldering van kleine moleculen zijn de laatste jaren steeds verder verbeterd, met een nadruk op het snel kunnen meten van verbindingen in grote series van monsters. Uit deze experimenten komen dan meestal moleculaire massa's of NMR pieken naar voren die, bijvoorbeeld, kenmerkend zijn voor een bepaalde plant variëteit of menselijke ziekte. Vaak is de exacte chemische structuur van de gemeten verbindingen (nog) onbekend en door de kleine hoeveelheid monster en verbinding niet eenvoudig te achterhalen.

Dit proefschrift heeft als doel de molecuulidentificatie te versnellen en voor lage concentratie verbindingen mogelijk te maken door de ontwikkeling van gecombineerde MS en NMR methodes. Dit kan met behulp van een systematisch gebruik van hoge resolutie MS, alsmede gedetailleerde MS fragmentatie ( $\text{MS}^n$ ), gecombineerd met 1-dimensionaal (1D)-NMR, en een geavanceerd NMR programma (PERCH), dat NMR data kan voorspellen, modelleren, en vergelijken. Deze combinatie maakt het mogelijk om op microgram niveau molecuulidentificatie uit te voeren. Deze studie laat zien dat het mogelijk is om de volledige structuur op te helderen van zeer opelkaar gelijkende fenolische verbindingen die geïsoleerd zijn uit complexe extracten van planten en menselijke urine. Het achterhalen van de structuren van deze verbindingen is van belang voor onderzoeken naar de functies van deze verbindingen en het leggen van relaties tussen de biobeschikbaarheid voor de mens van een verbinding en de structuur. Daarnaast wordt kwantificering van de geïdentificeerde verbindingen mogelijk, wat onderzoek naar de invloed van voedsel in het menselijk lichaam op moleculair niveau betrouwbaarder maakt.



In *Hoofdstuk 1* worden de achtergrond van het onderzoek en verschillende analysetechnieken en terminologieën die in dit proefschrift gebruikt worden toegelicht. *Hoofdstuk 2* beschrijft hoe een hoge resolutie MS<sup>n</sup> methode werd getest in zowel positieve als negatieve ionisatie modus om 121 fenolische verbindingen te kunnen onderscheiden en identificeren door structuurinformatie uit de fragmentatiepatronen te halen. De 121 referentieverbindingen werden met een injectie robot één voor één (offline) in de massaspectrometer gebracht. Deze aanpak resulteerde in reproduceerbare, robuuste en gedetailleerde MS<sup>n</sup> fragmentatie patronen, die zelfs het onderscheid tussen structureel zeer gelijkende verbindingen mogelijk maakte. Data van de accurate massa MS<sup>n</sup> spectrale patronen bleken een krachtig middel te zijn om moleculen met dezelfde elementformule (isomeren) te onderscheiden, wat essentieel is bij het identificeren en herkennen van verbindingen in complexe biologische monsters op basis van MS. In *hoofdstuk 3* is het gebruik van de MS<sup>n</sup> spectrale patronen methode getest bij het identificeren van fenolische verbindingen uit extracten van cherry tomaat (*Lycopersicon esculentum*) en de zandraket (*Arabidopsis thaliana*), een modelplant in de biologie. MS<sup>n</sup> spectrale patronen werden zowel direct na chromatografische scheiding (online) evenals met de injectie robot (offline) gegenereerd. De online gegenereerde MS<sup>n</sup> spectrale patronen waren minder uitgebreid door de beperkte meettijd die veroorzaakt wordt door de relatief korte tijd dat een verbinding na chromatografische scheiding in de MS zichtbaar is. Gedetailleerde MS<sup>n</sup> spectrale patronen van verbindingen uit de biologische extracten werden gegenereerd met behulp van een fractieverzamelaar/injectie robot. De resultaten van de twee MS<sup>n</sup> benaderingen bleken goed met elkaar overeen te komen. Met deze aanpak zijn 127 verbindingen geïdentificeerd, waarvan 21 moleculen niet eerder beschreven waren.

*Hoofdstuk 4* beschrijft de ontwikkeling van een geautomatiseerd vloeibare chromatografie (LC)-MS-vaste stof extractie (SPE)-NMR systeem dat in staat is om specifieke moleculen uit een complex biologisch extract te filteren op grond van hun massa, met als doel NMR spectra op te nemen van minimale hoeveelheden monster en verbinding. Tomaten waarvan de flavonoid biosynthese genetisch

sterk verhoogd was, werden gebruikt om de methode te testen. Deze combinatie van massa spectra en NMR signalen van de opgevangen verbindingen uit de LC-MS pieken leidde tot een efficiëntere opheldering van de chemische structuur van moleculen. Vervolgens was het mogelijk om het identificatieproces verder te versnellen met behulp van automatische 1D-<sup>1</sup>H-NMR spectrale voorspelling van experimentele data met de PERCH NMR software.

*Hoofdstuk 5* laat zien hoe de ontwikkelde analytische platformen gebruikt konden worden voor de succesvolle identificatie van 177 fenolische stoffen aanwezig in witte, groene, en zwarte thee. Hierbij zaten complexe geglycosyleerde verbindingen (moleculen geconjugeerd met een of meerdere suikergroepen), gerelateerd aan de flavonoïden kaempferol en quercetin, waarvan de chemische structuur, ondanks de zeer kleine ([sub-]microgram) hoeveelheden, toch volledig opgehelderd kon worden. De relatieve hoeveelheden van de fenolische verbindingen in 17 commercieel verkrijgbare zwarte, groene en witte thee producten werden gebruikt voor principale component analyse (PCA) om verschillen in de fenolische samenstelling tussen de thees te bestuderen. Hiermee konden de zwarte thee extracten onderscheiden worden van de groene en witte thee extracten, hetgeen aangeeft dat de fenolische samenstelling van zwarte thee afwijkt van die van groene en witte thee. Dit kan deels worden verklaard door de fermentatie van jonge (groene) thee bladeren tot zwarte thee, waarbij kleinere fenolische verbindingen polymeriseren tot grotere verbindingen die de bruine kleur van zwarte thee geven. De resultaten van deze studie laten zien dat het gecombineerd gebruik van MS<sup>n</sup> (LC-LTQ-Orbitrap FTMS) en NMR (LC-TOFMS-SPE-NMR) kan leiden tot een uitgebreidere en betrouwbaardere beschrijving van de verbindingen in thee en andere planten extracten.

In *Hoofdstuk 6* was het doel de uit thee afkomstige fenolische verbindingen aanwezig in de mens te identificeren en te kwantificeren door meting van stoffen die uitgescheiden worden in de urine. Daarvoor is een extra SPE concentratie stap gecombineerd met de twee aan hoge druk (HP)-LC gekoppelde analytische systemen zoals beschreven in hoofdstuk 3 en 4. Deze strategie heeft geresulteerd in de identificatie van 138 verbindingen in urine. Van 36 verbindingen is de

complete identiteit achterhaald, inclusief de conjugatiepatronen van diverse valerolacton verbindingen. Deze verbindingen zijn microbiële afbraakproducten van groene en zwarte thee verbindingen die niet zijn opgenomen in de dunne darm. NMR voorspellingen en modellen van NMR spectra van geglycuronideerde en gesulphateerde verbindingen zijn uitgevoerd ter controle van de NMR piektoekenningen, gebaseerd op alleen 1D-<sup>1</sup>H-NMR data, te bevestigen. Daarnaast zijn 26 uurs kwantitatieve uitscheidingsprofielen voor bepaalde geconjugeerde valerolactonen verkregen aan de hand van diagnostische NMR proton signalen in 1D-<sup>1</sup>H-NMR spectra van urine.

In *Hoofdstuk 7* worden de huidige status van molecuulidentificatie en de toekomstige uitdagingen wat betreft de volledige structuur opheldering van verbindingen met (sub)microgram hoeveelheden bediscussieerd. Het werk in dit proefschrift en de resultaten van andere groepen die aan de koppeling van MS en NMR werken laten zien dat de volledige identificatie van compleet nieuwe verbindingen op microgram niveau (en zelfs daaronder) mogelijk is door de combinatie van SPE trapping op basis van MS en 1D-<sup>1</sup>H-NMR of ultra-pressure LC (UPLC)-MS isolatie van verbindingen gevolgd door capillair NMR. De semi-automatische structuuropheldering van moleculen op basis van MS en NMR is nog volop in ontwikkeling en werkt al voor enkele moleculen en groepen van verbindingen. Om de automatisering van de molecuulidentificatie verder te ontwikkelen, zijn grotere databases nodig met hoge kwaliteit MS en NMR spectra, inclusief gebruiksvriendelijke software die kandidaatstructuren kan selecteren en verwerpen.

Alles tezamen draagt dit proefschrift bij aan de ontwikkeling van generieke methodes voor molecuulidentificatie met behulp van robuuste MS<sup>n</sup> fragmentatie spectra en 1D-<sup>1</sup>H-NMR. Daarnaast is spectrale data van goede kwaliteit verkregen van de geïdentificeerde verbindingen. Dit is van grote waarde voor verdere ontwikkelingen in het meten en analyseren van spectrale data en voor het ophelderen van nu nog onbekende verbindingen op een snellere, systematischere, en meer geautomatiseerde manier. Zo heeft deze aanpak in urine extracten

geleid tot een veel completer beeld van uit thee afkomstige verbindingen met onverwacht veel methyl (-CH<sub>3</sub>) geconjugeerde valerolacton en valeriaanzuur verbindingen.

De gevolgde molecuulidentificatie strategie kan worden gebruikt om ook in andere studies structuuropheldering te doen van verbindingen die biologische relevantie hebben. Met name bij studies met een beperkte hoeveelheid biologisch materiaal en lage concentraties aan relevante verbindingen is de gevolgde strategie erg nuttig. De kennis van de geïdentificeerde molecuulstructuren kan in vervolgstudies gebruikt worden. Tevens vormen de MS<sup>n</sup> spectrale patronen van de geïdentificeerde verbindingen een uitstekend begin voor een MS<sup>n</sup> spectrale database. Met de ontwikkelde protocollen is een dergelijke database snel uit te breiden met meerdere moleculen van verschillende groepen van verbindingen.

### *List of abbreviations used*

1D-<sup>1</sup>H-NMR, one dimensional proton nuclear magnetic resonance (spectroscopy);  
2D, two dimensional;  
3D, three dimensional;  
BT, black tea;  
C-AID, computer assisted identification;  
CID, collision induced dissociation;  
CRC, cross ring cleavage;  
DAD, diode array detector;  
Da, Dalton;  
EF, elemental formula;  
EI, electron impact;  
ESI, electrospray ionisation;  
FA, formic acid;  
FT, Fourier transformed;  
GAL, galactose moiety;  
GC, gas chromatography;  
GLC, glucose moiety;  
GT, green tea;  
HHDP, 6,6'-dicarbonyl-2,2',3,3',4,4'- hexahydroxybiphenyl moiety;  
HILIC, hydrophilic interaction liquid chromatography;  
HPLC, high pressure/performance liquid chromatography;  
LC, liquid chromatography;  
m/z, mass to charge;  
malonyl-CoA, malonyl coenzyme A;  
mDa, milliDalton;  
MeOD, deuterated methanol;  
MeOH, methanol;  
MS, mass spectrometry;

MS<sup>n</sup>, multi-stage fragmentation mass spectrometry;  
MI, metabolite identification;  
MSI, Metabolomics Standards Initiative;  
nA, nanoAmpère;  
NG/CNG, naringenin/chalconaringenin;  
NMR, nuclear magnetic resonance (spectroscopy);  
PCA, principal component analysis;  
SPE, solid phase extraction;  
TOF, time of flight;  
ref, reference;  
rf, radiofrequency;  
RO, ring opening;  
RT, retention time;  
RHM, rhamnose moiety;  
UPLC, ultra-high pressure/performance liquid chromatography;  
UV, ultraviolet;  
WT, white tea.







## *Acknowledgements*

As I am well aware that these pages are probably the most-read of my thesis, to get a hint of what I did during my PhD, I should refer you here to the summary of my thesis (page 229 [UK] and page 235 [NL])..(: But don't worry, you can also just enjoy the pictures throughout the booklet or read on here..(;

During the last 11 years I have been living in Wageningen and the last 4 years I performed my PhD study. I really enjoyed Wageningen as student city and as a place to live. Over the years, I explored more and more of Wageningen. During these discoveries, I met a lot of people that contributed to the nice atmosphere of Wageningen. As the space of the acknowledgements is rather limited I would like to thank all of you that made my time in Wageningen to a fantastic period of my life. Here I provide you some of the 'fragments' that made my life in Wageningen so beautiful during my PhD:

*Watching the sneak previews, Playing board games, Playing Floorball at WUV Stick Together and UFC Lions (actually in Rhenen!), Meeting study friends, Lively discussions during the lunch breaks, Mountainbiking around Wageningen (although 'Off-road' biking might be a better term..(;..), Shopping at the market, BBQ-ing at the Rhine, Enjoying the holidays-atmosphere in the garden of 't Binnenveld.....*

Of course, I will also use this space to give some personal words to people for their contribution to the fulfillment of my PhD grade.

Thanks a lot to all the colleagues of the NMC, Biochemistry, and Plant Research International for your advice, encouraging words, time for a coffee, or (just-in-time) being there. Piotr, Miguel, Julio, and Rob: thanks to your input the MS<sup>n</sup> methods were developed for plant compounds as well. Special thanks goes to Nico Nibbering and Albert Tas who commented on the proposed fragmentation pathways.

Laura, Jannie, Anneke, without your help.... well, I would be stuck with a huge pile of forms and have no nice booklet. Thanks a lot for all your support!! (:

Jonathan, Moktar, merci bien for being my MSc students. It was nice to have you over and see you growing as a scientist by planning and performing your experiments. I think you and I learnt a lot from it! And Moktar, 'merci bien' for your 'Tour de Bordeaux'..(:

Tomás, we spent quite some time together to get the machines up and running and to isolate your compounds from the flesh-eating plant extract and obtain useful NMR spectra, but in the end we managed to identify quite some compounds. And I am glad we could also meet in Portugal to see you ‘natural habitat’..(;

Yelda, thanks for your interest in the fragmentation data and I am glad I could make you enthusiastic for spectral analysis of the MS<sup>n</sup> data. You even started to like MS more than NMR.... I hope that you manage to find a suitable PhD project, and of course related to MS or NMR analysis.

Velitchka, thanks for all the advice during my PhD and I hope we can keep on collaborating on the NMR simulations, although we cannot yet predict the future very well with any software tool.... And I enjoyed our Canada trip through the Rockies after the Edmonton conference! (:

Lars, we had quite some lively discussions on how to combine the MS<sup>n</sup> data and your computational work in favor of metabolite annotation and this already resulted in a nice manuscript. I really hope we can continue this line of research.

Bert, thanks a lot for all the hours behind the machine and your support in contacting the vendor for yet another breakdown of a (vital!) electronic piece. I am quite sure that without your technical assistance this thesis would have taken me much longer to fulfill!

Sjef, your experience in LC-MS and capillaries was very useful since the LC-MS-SPE-NMR system heavily depends on..... exactly, lots of capillaries!

John, we had quite some lively discussions on the tea phenolic metabolites in the urine and how they (ac)cumulate in the urine, which did result in a nice article and chapter in my thesis. Thanks a lot for your input and advice.

Ric, thanks for all your assistance and advice during my PhD project. You returned my first manuscript texts nearly completely in red. Luckily I could more and more appreciate your useful comments and learnt how to express myself clearer. In the last manuscript text I even had to search for your comments! And the slides of the presentations were also thoroughly screened by you for pictures.... needless to say that I have a pile of good slides to show by now.

You were always willing to make time for me to explain and discuss on the running issues. Cheers to that with a (polyphenol-rich) tomato juice! (or do you prefer the red wine?) And did I already mention rutin? Hmm, I almost forgot!! (;

Jacques, you always stimulated me to follow my own path and take the initiative. Even though you were not convinced by the usefulness of the spectral tree approach in the beginning of the project, I was able to get you enthusiastic on the approach by the end of my PhD period. It is no secret that you also gave me some hard times during my PhD, but in the end we can be very satisfied with the obtained results. Cheers to that with a good glass of (polyphenol rich) red wine!

Sacco, you were there when I needed advice on how to continue my PhD work. Thanks a lot for that. And I am glad you kept your promise that one day we would fly over The Netherlands....it is a small world after all, especially as seen from the sky....

Raoul, thanks for taking care that I would finish nicely in time with a number of good papers. Our lively discussions helped me to explain myself clearer and stimulated me. Your enthusiasm should be an example for all scientists!

Uiteraard heb je ook thuis een goede basis nodig. Zo heb ik gedurende mijn hele PhD tijd met plezier in Centraal Wonen 't Binnenveld gewoond waar ik met mooi weer van de tuin heb kunnen genieten. En in die tuin heb ik ook Tineke mogen ontmoeten!

Bart, ook jij studeerde Moleculaire Wetenschappen en promoveerde in Wageningen. En daarnaast ging je ook nog eens unihockeyen. Dat maakt de kans wel erg groot dat je elkaar ontmoet. En zo geschiedde.... En toen ging je ook nog eens (samen met Marjon..(:..) op een steenworp afstand van mijn ouderlijk huis wonen! Hopelijk kunnen we nog vaak goede discussies voeren begeleid door enkele prima biertjes. Super dat je tijdens mijn promotie mijn paranimf wilt zijn.

Peter, jou inzet en brede interesse in de wetenschap zijn een voorbeeld voor iedere jonge wetenschapper. Jij bent je droom-PhD-baan in de VS gaan doen, waardoor we elkaar niet veel gezien hebben de afgelopen vier jaar. Maar voor jou geldt zeker: uit het oog is zeker niet uit het hart!! Alleen de volgende keer wordt het waarschijnlijk Amherst of Glasgow en niet Wageningen....  
Wie weet schrijven we ooit nog eens een ('serieus') artikel samen..(;

Harm, jij bent inmiddels net als ik een echte Wageninger geworden. We kennen elkaar al een tijdje en op het moment van schrijven ben je net getrouwd met Dori. Ik ben vereerd dat ik daar een speciale rol heb mogen vervullen en jullie te hebben zien stralen. De afgelopen jaren heb ik regelmatig geklaagd over de voortgang van het onderzoek, maar we hebben toch ook regelmatig het glas geheven op de behaalde successen. Dat heeft ertoe geleid dat mijn boekje vol is gekomen en mijn verdediging zomaar op 4 oktober gepland werd. Ik ben blij dat je op jou verjaardag mijn paranimf wilt zijn..(: En onze jaarlijkse fietstraditie zet zich gewoon voort in Schotland.., de Ardennen dit jaar waren een mooi oefenrondje..(;

Lieve Tineke, met onze plannetjes samen is elke dag weer een feest. Samen nieuwe gerechten uitproberen. Samen ontbijten met warme kampioentjes uit de oven. Het is gewoon fijn om bij jou te zijn. Hoe zo'n gevoel zo snel vertrouwd kan raken....(:

De laatste anderhalf jaar waren er regelmatig weekenden dat ik wat uurtjes 'achter de laptop' moest om zo dit boekwerk te vervolmaken. Ook zal ons leventje binnenkort op zijn kop gezet worden als ik richting Glasgow ga en ik je hopelijk met de mooie Highlands 'in de achtertuin' zoveel mogelijk tijd naar Schotland kan lokken..(; Ik zie uit naar mooie wandeltochten samen daar..(:  
Lieverd, I love you! (: (: Knuffieknuf van je beertje

De laatste woorden zijn voor mijn lieve ouders. Jullie hebben me altijd gesteund en gestimuleerd. In mijn jonge jaren hebben jullie vele uren met mij doorgebracht in het ziekenhuis, in wachtkamers, bij specialisten, maar zie: in 2001 ben ik als enthousiast eerstejaars naar Wageningen vertrokken en in 2012 nog altijd enthousiast als PhD richting Glasgow! Van 'Amersfoortse Hooglander' word ik dan 'Schotse Hooglander'..(:

Ik ben ontzettend blij dat jullie tijdens mijn jeugdige 'hoe werkt dat?' fase het alleen maar aangemoedigd hebben die vraag te blijven stellen, want dat blijft de basis van onderzoek doen. Lieve Hans en Truus, ik weet niet hoe ik jullie moet bedanken, dus doe ik het zo: ik hoop nog vele jaren met jullie onze vertrouwde uitjes naar Duitsland te kunnen maken en daar nu ook nieuwe plaatsen als Glasgow en Edinburgh aan toe te voegen! Een dikke knuffel van mij.



### *Curriculum Vitae*

Justin Johan Jozias van der Hooft was born on 4 September 1983 in Amersfoort, The Netherlands (NL). In 2001, he obtained his secondary education diploma at the Johan van Oldenbarneveltygymnasium in Amersfoort (NL). In September 2001, he started his BSc in Molecular Sciences at Wageningen University (NL), after which he continued with the MSc Molecular Sciences. During his master, he performed his two MSc theses at the Laboratory of Biophysics in Wageningen (NL), thereby studying a protein-DNA interaction with fluorescence techniques and the lycopene content of tomato by photoacoustic methods. In 2005-2006, Justin was member of the student co-management council of Wageningen University. To finalize his MSc, he did his internship in Copenhagen (Denmark) under the auspices of Prof. dr. Jerzy Jaroszewski at the Pharmaceutical Department of Copenhagen University. In Denmark, Justin analyzed South-African medicinal plants using NMR and MS technologies. It was there that he got interested in HPLC-SPE-NMR technologies. In September 2007, he received his MSc diploma in Wageningen, NL.

Starting in November 2007, Justin has worked for a couple of months as temporal co-worker at the Laborarotory of Biochemistry, Wageninen (NL), before he started his PhD in May 2008. Four years and one and a half month later he handed in his thesis of which at the time of the public defense all five experimental chapters have been successfully published in scientific journals.

Contact: [jjjvanderhooft at gmail.com](mailto:jjjvanderhooft@gmail.com)

## *Publication list*

### *First authorship:*

Van der Hooft, J. J. J., De Vos, R. C. H., Mihaleva, V., Bino, R. J., Ridder, L., de Roo, N., Jacobs, D. M., van Duynhoven, J. P. M., Vervoort, J., 2012a. Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake. *Anal. Chem.* 84, 7263-7271.

Van der Hooft, J. J. J., Akermi, M., Ünlü, F. Y., Mihaleva, V., Roldan, V. G., Bino, R. J., De Vos, R. C. H., Vervoort, J., 2012b. Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts. *J. Agr. Food Chem.* Article ASAP - DOI 10.1021/jf300297y.

Van der Hooft, J. J. J., Vervoort, J., Bino, R., de Vos, R.C.H., 2012c. Spectral trees as a robust annotation tool in LC-MS based metabolomics. *Metabolomics* 8, 691-703.

Van der Hooft, J. J. J., Mihaleva, V., Bino, R. J., de Vos, R. C. H., Vervoort, J., 2011a. A strategy for fast structural elucidation of metabolites in small volume plant extracts using automated MS-guided LC-MS-SPE-NMR. *Magn. Res. Chem.* 49, S55-S60.

Van der Hooft, J. J. J., Vervoort, J., Bino, R. J., Beekwilder, J., De Vos, R. C. H., 2011b. Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. *Anal. Chem.* 83, 409-416.

### *Co-authorship:*

Grevenstuk, T., Gonçalves, S., Domingos, T., Quintas, T., van der Hooft, J. J. J., Vervoort, J., Romano, A., 2012. Inhibitory activity of plumbagin produced by *Drosera intermedia* on food spoilage fungi. *Journal of the Science of Food and Agriculture* 92, 1638-1642.

Grevenstuk, T., van der Hooft, J. J. J., Vervoort, J., de Waard, P., Romano, A., 2009. Iridoid and caffeoyl phenylethanoid glycosides of the endangered carnivorous plant *Pinguicula lusitanica* L. (Lentibulariaceae). *Biochem. System. Ecol.* 37, 285-289.

Ridder, L., Justin J.J. van der Hooft, Stefan Verhoeven, Ric C.H. de Vos, René van Schaik, Vervoort, J., 2012. Substructure based annotation of high-resolution MS<sup>n</sup> spectral trees. In press.

Rojas-Cherto, M., Peironcely, J. E., Kasper, P. T., van der Hooft, J. J. J., De Vos, R. C. H., Vreeken, R. J., Hankemeier, T., Reijmers, T., 2012. Metabolite identification using automated comparison of high resolution MS<sup>n</sup> spectral trees. *Anal. Chem.* 84, 5524-5534.

Velasco, D. S., Baesso, M. L., Medina, A. N., Bicanic, D. D., Koehorst, R., Van Der Hooft, J. J. J., Bento, A. C., 2011. Thermal diffusivity of periderm from tomatoes of different maturity stages as determined by the concept of the frequency-domain open photoacoustic cell. *Journal of Applied Physics* 109, art. no. 034703.

*In preparation:*

Van der Hooft, J. J. J., De Vos, R. C. H., Mihaleva, V., Bino, R. J., Ridder, L., de Roo, N., Jacobs, D. M., van Duynhoven, J. P. M., Vervoort, J., 2012. The large scale identification and quantification of conjugates of intact and gut microbial bioconversion products of polyphenols.

Van Der Hooft, J. J. J., de Vos, R. C. H., Vervoort, J., Bino, R. J., 2012. Structural elucidation of metabolites at the micromolar level.

Beekwilder, J., Houwelingen, A., Romero de la Fuente, I., Capanoglu, E., Zamariola, L., de Vos, R. C. H., van der Hooft, J. J. J., van Beek, T., Boyacioglu, D., Orzaez, D., Mock, H.-P., Matros, A., Itkin, M., Aharoni, A., Bovy, A., Hall, R., 2012. Flavonoid glycoside formation and hydrolysis during tomato fruit tissue disruption.

Charnikhova, T. V., Kohlen, W., van der Hooft, J. J. J., Liu, W., Mulder, P. J., López Ráez, J., Geurts, R., Zwanenburg, B., Vervoort, J., Bouwmeester, H. J., 2012. Medicagol, a novel strigolactone with a unique structure occurring in *Medicago truncatula*.

## *List of co-author affiliations*

**Justin J.J. van der Hoof**

*Laboratory of Biochemistry, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

*Plant Research International, Wageningen University and Research Centre, POB 16, 6700 AA, Wageningen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

**Moktar Akermi**

*Laboratory of Biochemistry, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

**Jules Beekwilder**

*Plant Research International, Wageningen University and Research Centre, POB 16, 6700 AA, Wageningen, The Netherlands*

**Raoul J. Bino**

*Laboratory of Plant Physiology, Wageningen University, POB 658, 6700 AR, Wageningen, The Netherlands*

*Centre for Biosystems Genomics, POB 98, 6700 AB, Wageningen, The Netherlands*

*Agrotechnology and Food Sciences, Wageningen University, 6703 HD, Wageningen, The Netherlands*

**John P.M. van Duynhoven**

*Laboratory of Biophysics, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

*Unilever R&D, Vlaardingen, Olivier van Noortlaan 120, 3133 AT, Vlaardingen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

**Doris M. Jacobs**

*Unilever R&D, Vlaardingen, Olivier van Noortlaan 120, 3133 AT, Vlaardingen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*



## Velitchka Mihaleva

*Laboratory of Biochemistry, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

## Lars Ridder

*Laboratory of Biochemistry and Research Centre, Wageningen University, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

*Netherlands eScience Center, Science Park 140, 1098 XG Amsterdam, The Netherlands*

## Victoria Gomez Roldan

*Plant Research International, Wageningen University and Research Centre, POB 16, 6700 AA, Wageningen, The Netherlands*

*Netherlands Consortium for Systems Biology, P.O. Box 94215, 1090 GE, Amsterdam, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

## Niels de Roo

*Unilever R&D, Vlaardingen, Olivier van Noortlaan 120, 3133 AT, Vlaardingen, The Netherlands*

## Jacques Vervoort

*Laboratory of Biochemistry, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

## Ric C.H. de Vos

*Plant Research International, Wageningen University and Research Centre, POB 16, 6700 AA, Wageningen, The Netherlands*

*Centre for Biosystems Genomics, POB 98, 6700 AB, Wageningen, The Netherlands*

*Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC, Leiden, The Netherlands*

## Fatma Yelda Ünlü

*Laboratory of Biochemistry, Wageningen University and Research Centre, Dreijenlaan 3, 6703 HA, Wageningen, The Netherlands*



## *Overview of the completed training activities*

### Discipline Specific Activities

#### Courses and Trainings

Metabolomics Course, Leiden University, NL	2008
MicroTof Training, Bruker Bremen, Germany	2008
Orbitrap Training, Thermo, on location (PRI, Wageningen)	2008
Statistics of the omics, VLAG	2008

#### Conferences and Meetings

LC-NMR Conference, Max Planck Institute, Jena, Germany	2008
Metabolomics Society meeting Metab Society, Edmonton, Canada	2009
2 <sup>nd</sup> Polish Mass Spec. conf., Polish MS Society, Poznan, Poland	2010
Metabolomics2010, Metabolomics Society, Amsterdam, NL	2010
NuGoweek2011, Wageningen, NL	2011
NMR-DG, Delft, NL	2011
5 <sup>th</sup> International Conference on Polyphenols and Health (ICPH), Sitges, Spain	2011

#### General courses

Vlag PhD week	2008
Scientific writing in English, CENTA	2008
PhD Competence Assessment, WGS	2008
Project- and Time Management, WGS	2008
Scientific Publishing, WGS	2008
PCDI Post-doc retreat	2012

#### Educational

Teaching at practical courses of Practical Biological Chemistry (BIC)	2008-2012
---	-----------

#### Optional activities

Biochemistry biweekly seminars	2008-2012
NMC discussion days (including oral and poster presentations)	2008-2012
PhD retreat, Barcelona, Spain	2009

The research conducted in this thesis was financially supported by:  
The Netherlands Metabolomics Centre and  
The Centre for BioSystems Genomics,  
which are both part of:  
The Netherlands Genomics Initiative.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

This thesis is printed by GVO drukkers & vormgevers BV, Ede,  
The Netherlands.

Cover design and layout:  
Gideon van der Hooft - VANDERHOOFT PHOTOGRAPHY.  
All his efforts are gratefully acknowledged here.