

# Spectral trees as a robust annotation tool in LC–MS based metabolomics

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

Received: 26 June 2011 / Accepted: 8 September 2011 / Published online: 28 September 2011  
© Springer Science+Business Media, LLC 2011

**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>, using 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 using 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.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11306-011-0363-7) contains supplementary material, which is available to authorized users.

J. J. J. van der Hooft (✉) · J. Vervoort  
Laboratory of Biochemistry, Wageningen University  
and Research Centre, Dreijenlaan 3, 6703 HA Wageningen,  
The Netherlands  
e-mail: justin.vanderhooft@wur.nl

J. J. J. van der Hooft · R. C. H. de Vos  
Plant Research International, BU Bioscience,  
Wageningen University and Research Centre, P.O. Box 16,  
6700 AA Wageningen, The Netherlands

J. J. J. van der Hooft · J. Vervoort · R. C. H. de Vos  
Netherlands Metabolomics Centre, Einsteinweg 55,  
2333 CC Leiden, The Netherlands

R. J. Bino  
Laboratory of Plant Physiology, P.O. Box 658,  
6700 AR Wageningen, The Netherlands

R. J. Bino · R. C. H. de Vos  
Centre for BioSystems Genomics, P.O. Box 98,  
6700 AB Wageningen, The Netherlands

**Keywords** Annotation · Fragmentation · LC–MS ·  
Metabolomics · Metabolite identification · Structural  
elucidation

## 1 Introduction

The characterization, annotation and structural elucidation of metabolites in crude extracts are of great importance in MS based metabolomics (Bedair and Sumner 2008; Dunn 2008; Kind and Fiehn 2007; Moco et al. 2007a; Werner et al. 2008). 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 (Böttcher et al. 2008; Iijima et al. 2008; Moco et al. 2006; Sumner et al. 2007)). Tomato fruit peel has been subjected to metabolomics studies frequently (Gómez-Romero et al. 2010; Moco et al. 2006; Moco et al. 2007b; Vallverdú-Queralt et al. 2010). Likewise, leaves from *A. thaliana* plants have been frequently investigated using LC–MS (Böttcher et al. 2008; Matsuda et al. 2009; Roepenack-Lahaye et al. 2004; Tohge et al. 2007). Nevertheless, far from all metabolites have unambiguously been elucidated and thus completely annotated even in these important model species (Slimestad and Verheul 2009). In recent studies, the role of LC–MS and especially the need for MS fragmentation in the metabolite identification process is highlighted (Kind and Fiehn 2010; Rasche et al. 2011). 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 (Gómez-Romero et al. 2010). 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 (Vallverdú-Queralt et al. 2010). In addition, Ion Trap MS fragmentation up to MS<sup>4</sup> was used to detect 13 novel tomato seed compounds (Ferrerres et al. 2010). Moreover, recently a method was described (Francisco et al. 2009), based on Ion Trap MS fragmentation up to MS<sup>3</sup>, which enabled the simultaneous detection of >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 (Makarov et al. 2006; Makarov and Scigelova 2010) and proteomics (Scigelova and Makarov 2006), using the MS<sup>n</sup> fragmentation abilities of the Ion Trap in combination with the accurate mass provided by the Orbitrap FT–MS. 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, usually MS<sup>2</sup> and at maximum MS<sup>3</sup>, or a nominal mass resolution of MS<sup>n</sup> 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<sup>n</sup> fragmentation (Hooft et al. 2011) of metabolites resulting in so-called spectral trees (Sheldon et al. 2009). By using nanospray infusion into a LTQ/Orbitrap hybrid mass spectrometer, we could discriminate 119 out of 121 tested polyphenolic compounds including different series of isomers (Hooft et al. 2011). We concluded that the MS<sup>n</sup> 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<sup>n</sup> 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 (Hooft et al. 2011) 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<sup>n</sup> spectral trees to identify and discriminate flavonoid species in tomato and *Arabidopsis*, two plant species most frequently used in metabolomics 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 FT–MS was used to generate (partial) spectral tree data online during LC–MS analysis in an unbiased manner. Secondly, by using 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.

## 2 Materials and methods

### 2.1 Chemicals and plant material

#### 2.1.1 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.

#### 2.1.2 Plant harvesting and metabolite extraction

Tomato and *Arabidopsis* material were grown in the greenhouse and directly after harvesting, the plant material was frozen, ground, and stored at  $-80^{\circ}\text{C}$ . 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 (De Vos et al. 2007) with slight modifications (Supplemental Text S1 in the Supporting Information).

### 2.2 Analytical methods

#### 2.2.1 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 (Moco et al. 2006) (details in Supplemental Text S1; 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 FT-MS, respectively. MS settings and spectral tree topologies are given in Supplemental Text S1 (Supporting Information).

#### 2.2.2 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  $\mu\text{l}$ . The gradient and flow conditions were the same as described above, with an additional 30  $\mu\text{l}/\text{min}$  100% isopropanol added into the LC flow via a T-junction between the PDA and the NanoMate. The eluent flow was split by the NanoMate, at 219.5  $\mu\text{l}/\text{min}$  to the fraction collector and 0.5  $\mu\text{l}/\text{min}$  to the nano-electrospray source. LC-fractions

were collected every 5 s (i.e., 18  $\mu\text{l}$ ) into a 384 wells plate (Twin tec, Eppendorf), cooled at  $10^{\circ}\text{C}$ . After collection, 4  $\mu\text{l}$  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 (Supporting Information).

#### 2.2.3 Offline MS<sup>n</sup> spectral tree generation

The general analysis procedure and spectral tree topology have been described recently (Hooft et al. 2011). 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  $\mu\text{l}$  solvent from collected LC fractions, followed by direct infusion and nano-spraying into the MS (for details see Supplemental Text S1 in the Supporting Information).

### 2.3 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 one, 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 (Hooft et al. 2011).

Using 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 (Hooft et al. 2011). 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 FT-MS 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; in case of offline  $MS^n$  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 one and 70% in spectrum two, these two spectra were considered different.

#### 2.4 NMR identification of selected compounds

In order to verify compound annotations resulting from  $MS^n$  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-1H and 2D-COSY and 2D-TOCSY and 2D-HSQC NMR measurements on a 600 MHz NMR spectrometer (Bruker) (Moco et al. 2007a).

### 3 Results

Metabolites in crude plant extracts were separated using reversed-phase C18 analytical chromatography, followed by untargeted online LC- $MS^n$  in negative electrospray ionization mode. Parts of the chromatograms were subsequently subjected to LC-fractionation using the NanoMate robot (Fig. 1). Selected flavonoid glycosides, caffeoylquinic acids and, in case of *Arabidopsis*, glucosinolates (Fig. 2), were fragmented offline in both negative and positive ionization mode, and the resulting spectral data was listed (Supplemental Table S1 in the Supporting Information). The on the fly LC- $MS^n$  fragmentation of metabolites is referred to as online LC- $MS^n$ , whereas the  $MS^n$  fragmentation of LC-fractionated metabolites, using the NanoMate fractionation-injection robot, is referred to as offline  $MS^n$  fragmentation. The mass accuracy obtained during both online LC- $MS^n$  and offline  $MS^n$  analyses was always within 3 ppm, at all compound concentrations and  $MS^n$  levels, and for full scan ( $MS^1$ ) always within 1.5 ppm, in agreement with our previous results (Hooft et al. 2011).

#### 3.1 Robustness and reproducibility of online and offline $MS^n$ spectra

In order to determine the similarity between high mass resolution  $MS^n$  fragmentation patterns of metabolites and their fragments, the obtained  $MS^n$  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 (Hooft et al. 2011) and data obtained within the present study (see Sect. 2.3).

Multiple online LC- $MS^n$  spectra from the same compound can be obtained by continuously generating spectra from its

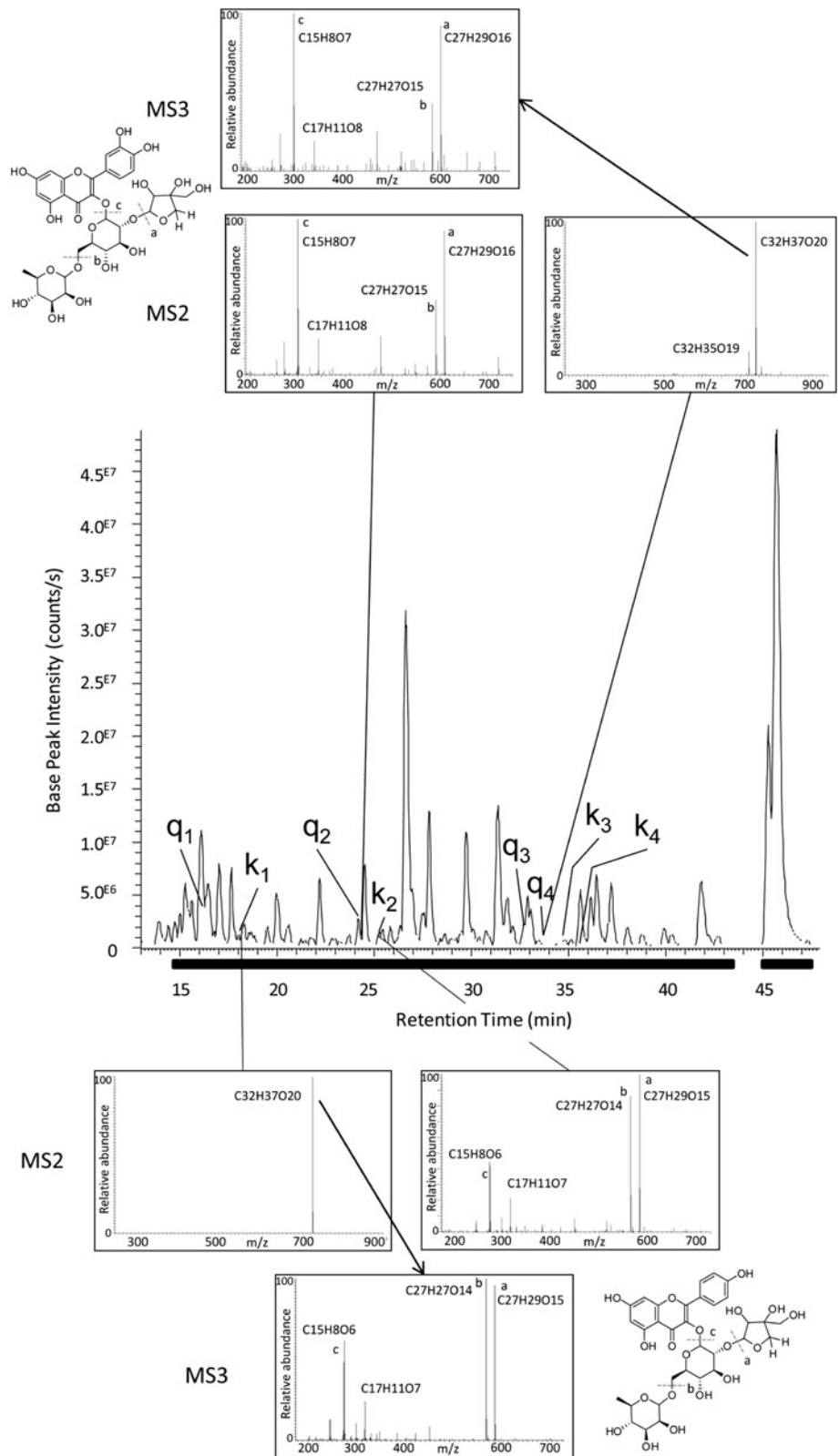
eluting chromatographic peak. This online spectra generation implies that metabolite spectra are obtained at different parent ion concentrations. Rutin (quercetin-3-*O*-rutinoside), a commonly occurring flavonoid-glycoside in plants, was taken as an example to test the reproducibility of online LC- $MS^n$  spectra. During chromatography of tomato fruit, four repetitive LC- $MS^3$  spectra of the quercetin fragment of rutin were obtained (Fig. 3a). While the parent ion intensities ranged from  $2E6$  to  $4E7$ , the differences in relative intensities of all fragment ions present in the  $MS^3$  spectra were always within 20% difference and in most cases differing less than 5%. The robustness of the online LC- $MS^n$  approach was further tested by comparing  $MS^3$  spectra of the chromatographic peak of quercetin-3-*O*-(2''-apiofuranosyl-6''-rhamnosylglucoside), present in tomato fruit peel, obtained *in duplo* and with a 6 month time interval (Fig. 3b). Even though the absolute intensities of the fragment ions varied slightly, the resulting four  $MS^3$  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  $MS^n$  approach, as well as with those of its authentic standard measured by offline direct infusion (Hooft et al. 2011) (Fig. 3c). The fragmentation spectra were all identical. These results illustrate the robustness and reproducibility of the online high mass resolution LC- $MS^n$  approach when using the normalized fragmentation spectra for comparisons.

#### 3.2 Online LC- $MS^n$ and offline $MS^n$ spectral trees of crude plant extracts

Since the online high mass resolution LC- $MS^n$  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  $MS^n$  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  $MS^n$  spectral data from their respective reference compounds (Hooft et al. 2011). The resulting fragments of the analyzed metabolites from tomato and *Arabidopsis* are provided in Supplemental Table S1 (Supporting Information).

The online LC- $MS^n$  approach resulted in 270  $MS^2$  scan events in the Orbitrap for a specific  $m/z$  value. As they are multiple series of isomers present in the extract, which results for some  $m/z$  values in multiple scan events, the number of unique  $MS^2$  scans is estimated to be around 450. In  $MS^3$ , the number of scan events per unique  $m/z$  is 490, leading to an

**Fig. 1** Part of a tomato fruit LC-Orbitrap FT-MS chromatogram in negative ionization mode. *Black bars* indicate the chromatographic sections fractionated each 5 s 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. Online 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 online 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*



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.

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

**Fig. 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

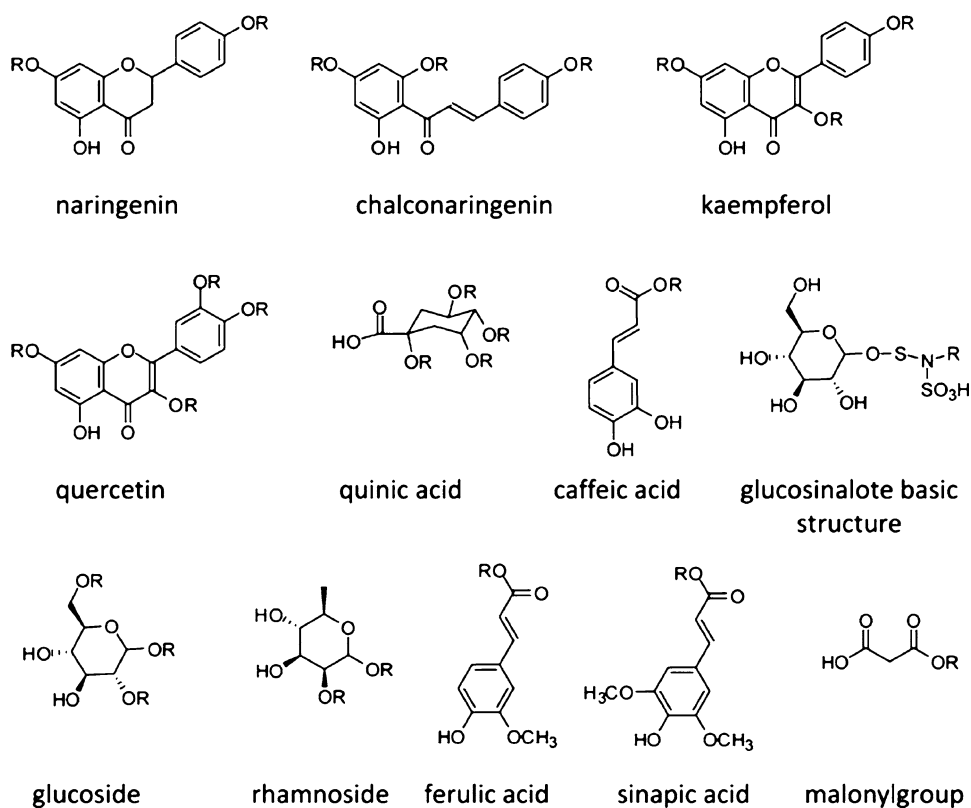


Table S1 in the Supporting Information). Although most of the measured metabolite intensities (1E5–5E6 counts/s) were lower than those used for the reference compounds [1E7–1E8 counts/s, (Hooft et al. 2011)], detailed fragmentation spectra ranging from MS<sup>3</sup> up to MS<sup>5</sup>, depending on compound concentration and ionization efficiency of fragments, could be obtained for almost all flavonoids. In contrast, on-line 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 four quercetin (q1–q4) and four kaempferol (k1–k4) conjugates detected in tomato peel (Fig. 1), their three most complex glycosides (q1, q3, q4, and k1, k3, k4) produced MS<sup>3</sup> spectra that were similar to the MS<sup>2</sup> spectra of their lesser conjugated analogues (q2 and k2, respectively, structures shown in Fig. 1). This result shows the ability of MS<sup>n</sup> to elucidate common substructures within related complex metabolites.

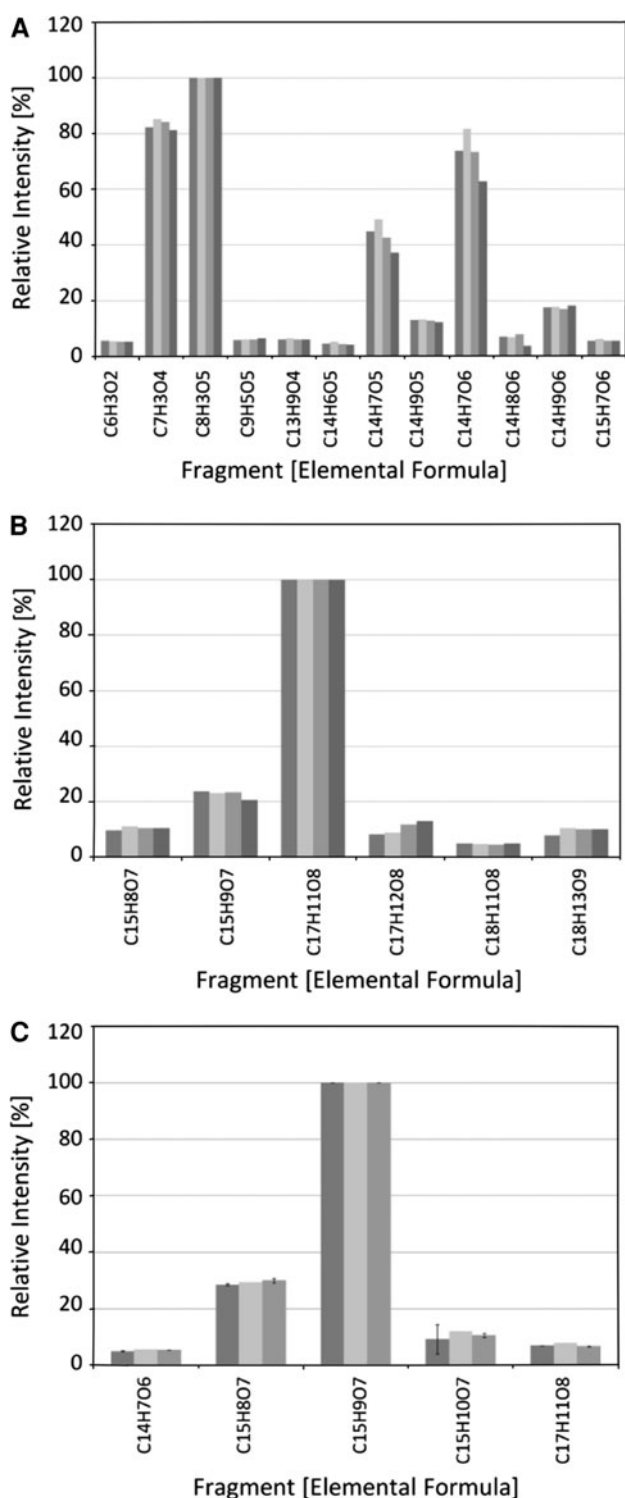
### 3.3 Effects of dynamic exclusion and mass resolution in online LC–MS<sup>n</sup>

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 s 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 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 FT–MS. 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 (1,050) 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



**Fig. 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''-rhamnosyl)glucoside 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

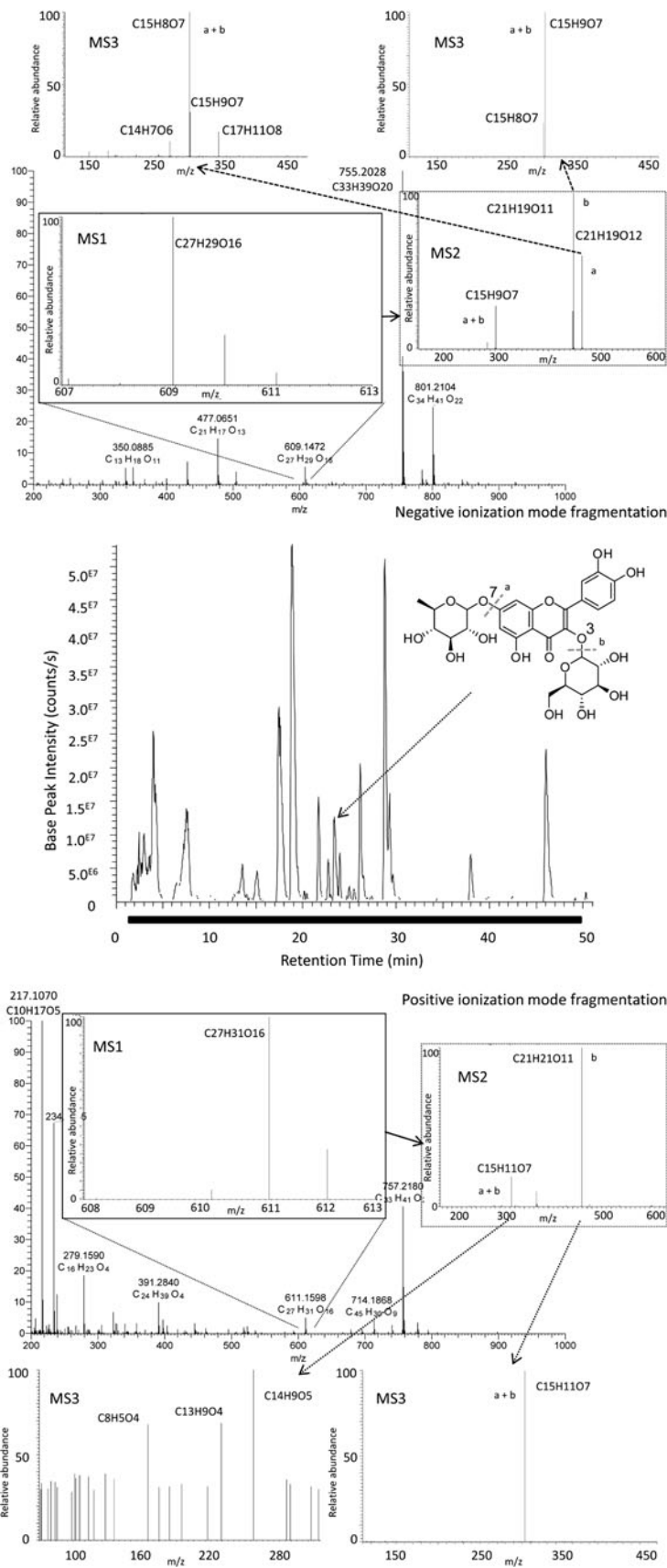
MS<sup>2</sup> and the MS<sup>3</sup> spectra were not identical between the two analyzers, as two minor fragments were only detected in the Orbitrap spectra. Nevertheless, the variations in relative intensities of the fragments detected by both analyzers were well within the criterion of 20%. Therefore, we conclude that spectral trees generated by the Orbitrap and the Ion Trap match well.

#### 3.4 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 min 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 MS<sup>1</sup> 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 (Hoofst et al. 2011), most structural information was retrieved using negative mode spectra. Based on the differences in the two negative mode MS<sup>3</sup> spectra displayed in Fig. 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 Fig. 4 for structure). Also in tomato, the fragmentation in both ionization modes was key to discriminate between some

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

**Fig. 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





isomers (e.g., Supplemental Figs. 5 and 6), thus illustrating the added value of performing offline MS<sup>n</sup> fragmentation.

### 3.5 MS<sup>n</sup> spectra enable discrimination of isomers

During full scan LC-Orbitrap FT-MS 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 Fig. 7. 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 Fig. 1E) 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 Fig. 2) and positive ionization mode (Supplemental Fig. 3), and based on the similarity with C<sub>15</sub>H<sub>12</sub>O<sub>5</sub> standards (Hooft et al. 2011) led 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 seven selected isomers of phenylpropanoids (Supplemental Fig. 4) and flavonoids (Supplemental Figs. 5–7) in the crude tomato extract.

### 3.6 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 compounds, and used an extract of *Arabidopsis* leaves to validate the MS<sup>n</sup> approach as generic tool. Supplemental Table S1 (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 (Iijima et al. 2008), while for 47 compounds only limited information has yet been provided, usually an elemental formula and retention time (Gómez-Romero et al. 2010; Iijima et al. 2008; Moco et al. 2006), 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 Fig. 8. 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 Supplementary Table S3). 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 four out of the five detected C<sub>30</sub>H<sub>28</sub>O<sub>13</sub> isomers have been described before in tomato (Iijima et al. 2008), 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 Fig. 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 (Hooft et al. 2011). The NMR measurements indeed identified these two metabolites as kaempferol-3-*O*-(2-rhamnosylglucoside)-7-*O*-rhamnoside and kaempferol-3-*O*-(6-glucosylglucoside)-7-*O*-rhamnoside. Thus, these four examples show that the structural prediction of metabolites based on MS<sup>n</sup> patterns was confirmed by unambiguous structural elucidation using NMR.

## 4 Discussion

LC-MS based metabolomics approaches are widely used for profiling of complex biological extracts such as from

plants (Dunn 2008; Werner et al. 2008). 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 S3 and (Kind and Fiehn 2010)]. 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, by using 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 (Iijima et al. 2008; Kind and Fiehn 2010; Moco et al. 2006). 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 (Fig. 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 (Fig. 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 (Fig. 3a) and from dilution series of standards, and MS<sup>n</sup> spectra obtained at different normalized collision energies of the Ion Trap (Hooft et al. 2011), and over prolonged time between spectral tree generations [Fig. 3b and (Hooft et al. 2011)]. 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 (Hooft et al. 2011), 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).

#### 4.1 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 spectral 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 Fig. 1A–D). 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 using accurate mass (Supplemental Figs. 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.

#### 4.2 Metabolite discrimination and annotation using MS<sup>n</sup> fragmentation

Polyphenols (Vukics and Guttman 2010; Wolfender et al. 2000) and glucosinolates (Millán et al. 2009; Rochfort

et al. 2008) 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 (Hooft et al. 2011), 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 (Olsen et al. 2009; Rochfort et al. 2006; Vukics and Guttman 2010) and glucosinolates (Rochfort et al. 2008), 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, by using 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 Figs. 1–7 and c.f. (Clifford et al. 2005)]. 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 (Supp. Table 1). 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 (Olsen et al. 2009; Rochfort et al. 2006). For instance, our fragmentation of the intact glucosinolates in *Arabidopsis* gave similar results compared to earlier Ion Trap fragmentation studies (Millán et al. 2009; Rochfort et al. 2008), indicating the reproducibility of MS<sup>n</sup> spectra even on different Ion Trap based mass spectrometry platforms. The examples of substitutions on flavonoids (Figs. 1 and 4) demonstrate that, in negative ionization mode, substitutions on the 7-O position fragmented more easily and generated different spectral trees as compared to 3-O substitutions (Hooft et al. 2011; Ma et al. 2001). 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 annotation in practice. In addition, previously identified metabolites can be easily detected as substructures within more complex compounds (Fig. 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 s 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<sup>n</sup> spectra in an unbiased manner. Although the spectral tree concept as suggested by Sheldon et al. is quite similar to ours, we used accurate mass instead of nominal mass MS<sup>n</sup> data, focused on its reproducibility aspects and applied it on a larger set of reference compounds (Hooft et al. 2011; Sheldon et al. 2009) and tested and optimized its applicability for annotating yet known and unknown compounds in crude extracts (present study). In addition,

while the MassFrontier software can visualize spectral trees (Sheldon et al. 2009), it currently has substantial drawbacks, like the inclusion of noise peaks due to the fact that masses instead of elemental formulas are used (Hooft et al. 2011). 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<sup>n</sup> spectral tree data, as well as spectral tree viewing and library searching, which are key steps in automation of metabolite identification based on MS<sup>n</sup> 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<sup>n</sup> spectra (Rasche et al. 2011), 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<sup>n</sup> data will very likely become available in the nearby future. Once this software is available, MS<sup>n</sup> 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<sup>n</sup> patterns from metabolites in complex crude extracts needed as input for this dedicated MS<sup>n</sup> software tools and databases.

## 5 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> spectral trees is foreseen in the nearby future.

**Acknowledgments** 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, Thomas Hankemeier, 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.

## References

- Bedair, M., & Sumner, L. W. (2008). Current and emerging mass-spectrometry technologies for metabolomics. *TrAC - Trends in Analytical Chemistry*, 27, 238–250.
- Böttcher, C., Roepenack-Lahaye, E. V., Schmidt, J., Schmotz, C., Neumann, S., Scheel, D., et al. (2008). Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in *Arabidopsis*. *Plant Physiology*, 147, 2107–2110.
- Clifford, M. N., Knight, S., & Kuhnert, N. (2005). Discriminating between the six isomers of dicaffeoylquinic acid by LC–MS<sup>n</sup>. *Journal of Agricultural and Food Chemistry*, 53, 3821–3832.
- De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 2, 778–791.
- Dunn, W.B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Physical Biology*, 5(1), 011001.
- Ferreres, F., Taveira, M., Pereira, D. M., Valentão, P., & Andrade, P. B. (2010). Tomato (*Lycopersicon esculentum*) seeds: New flavonols and cytotoxic effect. *Journal of Agricultural and Food Chemistry*, 58, 2854–2861.
- Francisco, M., Moreno, D. A., Cartea, M. E., Ferreres, F., García-Viguera, C., & Velasco, P. (2009). Simultaneous identification of glucosinolates and phenolic compounds in a representative collection of vegetable *Brassica rapa*. *Journal of Chromatography A*, 1216, 6611–6619.
- Gómez-Romero, M., Segura-Carretero, A., & Fernández-Gutiérrez, A. (2010). Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. *Phytochemistry*, 71, 1848–1864.
- Hooft, J. J. J.v.d., Vervoort, J., Bino, R. J., Beekwilder, J., & Vos, C. H. R.d. (2011). Polyphenol identification based on systematic

- and robust high-resolution accurate mass spectrometry fragmentation. *Analytical Chemistry*, 83, 409–416.
- Iijima, Y., Suda, K., Suzuki, T., Aoki, K., & Shibata, D. (2008). Metabolite profiling of chalcones and flavanones in tomato fruit. *Journal of the Japanese Society for Horticultural Science*, 77, 94–102.
- Kind, T.M., & Fiehn, O. (2007). Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, 8, 105.
- Kind, T., & Fiehn, O. (2010). Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical Reviews*, 2, 23–60.
- Ma, Y. L., Cuyckens, F., Van den Heuvel, H., & Claeys, M. (2001). Mass spectrometric methods for the characterisation and differentiation of isomeric O-diglycosyl flavonoids. *Phytochemical Analysis*, 12, 159–165.
- Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., et al. (2006). Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Analytical Chemistry*, 78, 2113–2120.
- Makarov, A., & Scigelova, M. (2010). Coupling liquid chromatography to Orbitrap mass spectrometry. *Journal of Chromatography A*, 1217, 3938–3945.
- Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., & Saito, K. (2009). MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant Journal*, 57, 555–577.
- Millán, S., Sampedro, M. C., Gallejones, P., Castellón, A., Ibargoitia, M. L., Goicolea, M. A., et al. (2009). Identification and quantification of glucosinolates in rapeseed using liquid chromatography–ion trap mass spectrometry. *Analytical and Bioanalytical Chemistry*, 394, 1661–1669.
- Moco, S., Bino, R. J., De Vos, R. C. H., & Vervoort, J. (2007a). Metabolomics technologies and metabolite identification. *TrAC - Trends in Analytical Chemistry*, 26, 855–866.
- Moco, S., Bino, R. J., Vorst, O., Verhoeven, H. A., de Groot, J., van Beek, T. A., et al. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiology*, 141, 1205–1218.
- Moco, S., Capanoglu, E., Tikunov, Y., Bino, R. J., Boyacioglu, D., Hall, R. D., et al. (2007b). Tissue specialization at the metabolite level is perceived during the development of tomato fruit. *The Journal of Experimental Botany*, 58, 4131–4146.
- Olsen, H., Aaby, K., & Borge, G. I. A. (2009). Characterization and quantification of flavonoids and hydroxycinnamic acids in curly kale (*Brassica oleracea* L. convar. *acephala* var. *sabellica*) by HPLC-DAD-ESI-MS<sup>n</sup>. *Journal of Agricultural and Food Chemistry*, 57, 2816–2825.
- Rasche, F., Svatos, A., Maddula, R. K., Böttcher, C., & Böcker, S. (2011). Computing fragmentation trees from tandem mass spectrometry data. *Analytical Chemistry*, 83, 1243–1251.
- Rochfort, S. J., Imsic, M., Jones, R., Trenerry, V. C., & Tomkins, B. (2006). Characterization of flavonol conjugates in immature leaves of pak choy [*Brassica rapa* L. Ssp. *chinensis* L. (Hanelt.)] by HPLC-DAD and LC-MS/MS. *Journal of Agricultural and Food Chemistry*, 54, 4855–4860.
- Rochfort, S. J., Trenerry, V. C., Imsic, M., Panozzo, J., & Jones, R. (2008). Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MS<sup>n</sup> fragmentation. *Phytochemistry*, 69, 1671–1679.
- Roepenack-Lahaye, E.v., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., et al. (2004). Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiology*, 134, 548–559.
- Scigelova, M., & Makarov, A. (2006). Orbitrap mass analyzer—overview and applications in proteomics. *Proteomics*, 1, 16–21.
- Sheldon, M. T., Mistrik, R., & Croleya, T. R. (2009). Determination of ion structures in structurally related compounds using precursor ion fingerprinting. *Journal of the American Society for Mass Spectrometry*, 20, 370–376.
- Slimestad, R., & Verheul, M. (2009). Review of flavonoids and other phenolics from fruits of different tomato *Lycopersicon esculentum* Mill. cultivars. *Journal of Science Food Agriculture*, 89, 1255–1270.
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, 3, 211–221.
- Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., & Saito, K. (2007). Phytochemical genomics in *Arabidopsis thaliana*: A case study for functional identification of flavonoid biosynthesis genes. *Pure and Applied Chemistry*, 79, 811–823.
- Vallverdú-Queralt, A., Jáuregui, O., Medina-Remón, A., Andrés-Lacueva, C., & Lamuela-Raventós, R. M. (2010). Improved characterization of tomato polyphenols using liquid chromatography/electrospray ionization linear ion trap quadrupole Orbitrap mass spectrometry and liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 24, 2986–2992.
- Vukics, V., & Guttman, A. (2010). Structural characterization of flavonoid glycosides by multi-stage mass spectrometry. *Mass Spectrometry Reviews*, 29, 1–12.
- Werner, E., Heilier, J. F., Ducruix, C., Ezan, E., Junot, C., & Tabet, J. C. (2008). Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *Journal of Chromatography B*, 871, 143–163.
- Wolfender, J. L., Waridel, P., Ndjoko, K., Hobby, K. R., Major, H. J., & Hostettmann, K. (2000). Evaluation of Q-TOF-MS/MS and multiple stage IT-MS<sup>n</sup> for the dereplication of flavonoids and related compounds in crude plant extracts. *Analysis*, 28, 895–906.