ORIGINAL ARTICLE

Liquid chromatography–mass spectrometry-based metabolomics for authenticity assessment of fruit juices

Lukas Vaclavik · Andre Schreiber ·
Ondrej Lacina · Tomas Cajka · Jana Hajslova

Abstract High performance liquid chromatography–mass spectrometry (HPLC–MS) technique, employing a hybrid triple quadrupole/linear ion trap (QqQ/LIT) mass analyzer, was used for comprehensive metabolomic fingerprinting of several fruit juices types, prepared from expensive (orange) or relatively low-priced (apple, grapefruit) fruits. Following the automated data mining and pre-treatment step, the suitability of the multivariate HPLC–MS metabolomic data for authentication, i.e., classification of fruit juice and adulteration detection, was assessed with the use of advanced chemometric tools (principal component analysis, PCA, and linear discrimination analysis, LDA). The LDA classification model, constructed and validated employing a highly variable samples set, was able to reliably detect 15% addition of apple or grapefruit juice to orange juice. In the final stage of this study, high performance liquid chromatography–quadrupole–quadrupole-time-of-flight mass spectrometry (HPLC–QqTOFMS) measurements were performed in order to obtain data for identification of pre-selected marker compounds using elemental formula calculation and online databases search.

Keywords Fruit juices · Metabolomics · Authenticity · Adulteration · Liquid chromatography · Mass spectrometry

L. Vaclavik · O. Lacina · T. Cajka · J. Hajslova (✉)
Department of Food Chemistry and Analysis, Institute of
Chemical Technology, Prague, Faculty of Food and Biochemical
Technology, Technicka 3, 16628 Prague 6, Czech Republic
e-mail: jana.hajslova@vscht.cz

A. Schreiber
AB SCIEX, 71 Four Valley Drive, Concord, ON L4K 4V8,
Canada

1 Introduction

The production of fruit juices represents an important and rapidly growing sector of beverage industry. Beside of orange juice, which is produced and consumed in the largest volume worldwide, also other fruit juice types, such as those obtained from pomegranate or various types of berries have become popular with regard to high levels of antioxidants presumed to be associated with positive health effects. Similarly to other highly prized food commodities, the economic value and large-scale production of orange juice and other valuable fruit juices have made them a likely target for adulteration and fraud (Ebeler et al. 2007). The most frequent profit-driven fraudulent procedures applied either alone or in combination, employ dilution with water, addition of sugars or pulp wash, and extension of authentic juice with cheaper alternatives (typically juices obtained from apples, pears, grapefruits, etc.) (Muntean 2010).

As the adulteration of fruit juices represents an ongoing problem, suitable analytical methods are needed to control authenticity parameters dictated by the legislation (Council Directive 2001/112/EC 2001). Until now, a number of methods have been developed to tackle various aspects of fruit juice authenticity (Fügel et al. 2005). The most established approaches are based on profiling of carbohydrates, phenolic compounds, carotenoids, amino or other organic acids using chromatographic methods such as high performance liquid chromatography (HPLC) or gas chromatography (GC) coupled to various types of detectors (Gómez-Ariza et al. 2005; Ehling and Cole 2011; Muntean 2010; Catillo et al. 2003; Pan et al. 2002; Low 1996). Additionally, electromigration (Saavedra et al. 2000; Saavedra et al. 2001; Jezek and Suhaj 2001) and spectroscopic methods (Cuny et al. 2008; Le Gall et al. 2001;

Springer

Kelly and Downey 2005; He et al. 2007) have been employed for this purpose. Another strategy enabling fruit juice authentication is based on examination of volatile components with the use of GC (Tønder et al. 1998; Shaw et al. 1999) technique. With regard to targeted nature of all the above mentioned methods, each of them can be successfully applied to monitor one or only few specific adulteration practices. It should be noted that fraud performers are usually one step ahead of the available testing methods, as new and more sophisticated adulteration methods are continuously developed. Therefore, analytical approaches and platforms facilitating more comprehensive insight into chemical composition of fruit juices and its changes, as associated with adulteration, are required.

The field of metabolomics, aiming at global analysis of numerous targeted or non-targeted low molecular compounds (metabolites) in a biological sample, has recently found its application in diverse research areas (Kaddurah-Daouk and Krishnan 2008; Wishart 2008) including food quality and authenticity assessment (Cevallos–Cevallos et al. 2009). Rapid growth of metabolomics has been enabled by substantial advances in analytical techniques such as mass spectrometry (MS) coupled to LC, GC, or CE and nuclear magnetic resonance (NMR), all techniques facilitating analysis of a wide range of metabolites with diverse physico-chemical properties, occurring at different concentration levels (Dunn and Ellis 2005). To process and interpret complex data obtained within metabolomic-based studies, advanced data mining/data processing software algorithms and multivariate chemometric tools are needed. Principal component analysis (PCA) represents the most frequently employed method for initial exploration of data internal structure and sample clustering. In the next step, supervised pattern recognition methods, such as linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA), support vector machine (SVM), or artificial neural networks (ANNs) are used for construction of classification and prediction models (Berrueta et al. 2007).

In this study, the feasibility of HPLC–MS technique employing a hybrid triple quadrupole/linear ion trap (QqQ/LIT) mass analyzer for metabolomic-based authentication of fruit juices (orange, grapefruit, apple, and their admixtures) representing different price categories was explored. Highly complex HPLC–MS records were treated using an automated data mining algorithm and further processed with multivariate chemometric methods (PCA and LDA) to asses the suitability of the data to differentiate respective juice types and to detect their adulteration. In the next step, accurate single MS and MS/MS mass spectra were acquired for the most characteristic markers using quadrupole–quadrupole-time-of-flight mass spectrometry (QqTOFMS) to estimate their probable elemental formulae and enable tentative identification.

2 Experimental

2.1 Chemicals

Deionized water used for both as LC mobile phase and for juice samples dilution was obtained from a Milli-Q purification system (Millipore, Bedford, MA, USA). HPLC-grade methanol and ammonium acetate (purity ≥99%) were supplied by Sigma–Aldrich (St. Luis, MO, USA).

2.2 Samples and sample preparation

Fruit juices obtained from orange ($n = 19$), apple ($n = 16$), and grapefruit ($n = 16$) were purchased from reliable sources at the Czech and Canadian retail markets. Collected samples represented both fresh-pressed juices ($n = 8$) and juices prepared from concentrate ($n = 43$) and were produced in various countries (Austria, Czech Republic, Hungary, Poland, Romania, Slovakia, Spain, and USA). To simulate adulteration, admixtures of expensive and cheap fruit juices were prepared in volume ratios specified in Table 1. Fruit juice samples (10 ml) were centrifuged (5 min, 10,000 rpm, 20°C) to remove solid particles; subsequently an aliquot volume (2 ml) was passed through a 0.22 µm PTFE membrane filter, 100-fold diluted with deionized water and transferred into an auto-sampler vial. For the purpose of HPLC–QqTOFMS analyses, additional dilution (tenfold) of selected samples was carried out. Samples were stored in dark and dry at 2°C, until LC–MS measurements.

2.3 HPLC–MS

Two MS instruments using the same Agilent Series 1200 LC system (consisting of a vacuum degasser, an auto-sampler, a binary pump, and a thermostatted column compartment, Agilent Technologies, Santa Clara, CA, USA) were employed in this study. While an AB SCIEX 4000 QTRAP hybrid triple quadrupole/linear ion trap mass spectrometer (AB SCIEX, Concord, Ontario, Canada) was used for the measurement of sample metabolomic fingerprints, experiments aimed at marker compounds identification were performed with an AB SCIEX TripleTOF 5600 quadrupole-time-of-flight mass spectrometer (AB SCIEX, Concord, ON, Canada).

Chromatographic separation was performed using a Restek Ultra II Aqueous C18 reversed phase analytical column (50 × 2.1 mm i.d., 3 µm particle size, Restek, Bellefonte, Pennsylvania, USA) kept at a temperature of 35°C. The injected sample volume was 10 µl. The mobile phase consisted of 5 mM ammonium acetate (eluent A) and methanol (eluent B). The gradient elution was performed as follows: 0–3 min eluent B 10%; 3–5 min eluent

Table 1 Description of in laboratory prepared fruit juices admixtures

Admixture		Composition ratios (A:B, % v/v)	Number of individual admixtures at each ratio[a]
Component A	Component B		
Orange	Apple	90:10; 85:15, 75:25; 50:50; 25:75	3
Orange	Grapefruit	90:10; 85:15; 75:25; 50:50; 25:75; 10:90	3

[a] Different randomly selected juices were used for preparation of respective admixtures

B 10–90%; 5–7.5 min eluent B 90%; 7.5–10 min column equilibration-eluent B 10%. A flow rate of 0.4 ml min^{-1} was employed for elution; equilibration of the column was carried out at 0.6 ml min^{-1}.

A 4000 QTRAP was equipped with a Turbo V electrospray (ESI) ion source, whereas a TripleTOF 5600 utilized a Duo Spray source with separated ESI and atmospheric-pressure chemical ionization (APCI) probes. The ESI was used for the measurement of sample metabolomic fingerprints, while APCI probe worked as the second gas heater; APCI was also utilized for exact mass calibration of the QqTOFMS instrument. With regard to similarity of respective ion sources, identical parameter settings were used. In the positive ESI mode parameters were as follows: capillary voltage: +5000 V, nebulizing gas pressure: 50 psi, drying gas pressure: 60 psi, temperature: 500°C, and declustering potential: 60 V. The capillary voltage in negative ESI was −4000 V, other source settings were the same as for positive ESI.

A 4000 QTRAP was operated using an information dependent acquisition method to collect full scan MS and MS/MS information simultaneously. The two methods (positive and negative polarity) consisted of a survey scan using Enhanced MS scanning (EMS) from m/z 100 to m/z 1000 at 4000 amu s^{-1} and Enhanced Product Ion (EPI) scanning from m/z 50 to m/z 1000 at 4000 amu s^{-1}. EPI spectra were collected automatically for the two most intense ions in the EMS scan, above a threshold of 3000 counts per second (cps), throughout the chromatographic run. Dynamic background subtraction was activated to automatically acquire EPI spectra of co-eluting compounds. EPI spectra were collected using a collision energy of 35 V with a collision energy spread of ±15 V. Collision energy spread resulted in more characteristic MS/MS spectra since both low and high energy fragment ions were present in a single spectrum. Dynamic fill time was activated in EMS and EPI to avoid possible space charge for compounds at higher concentration.

Similar experiments, with identical IDA settings as on QTRAP system, were performed using TripleTOF 5600. Due to principle of QqTOF mass spectrometer, the collision voltage was linearly ramped from 20 to 50 V, instead of MS/MS spectra acquisition under three different discreet collision voltages settings. The high resolution survey spectra (mass resolving power >30,000 fwhm) were collected for 100 ms, subsequently, 10 ions were selected for acquiring MS/MS fragmentation spectra acquisition (each collected for 50 ms) after each scan. The exact mass calibration was performed automatically before each analysis employing the Automated Calibration Delivery System.

It should be noted that the in-batch sequence of the samples was random (established based on random number generation) to avoid any possible time dependent changes in HPLC–MS chromatographic fingerprints. Within each sequence run, blank samples (deionized water) were analyzed.

2.4 Data processing and chemometric analysis

MarkerView software (version 1.2.1, AB SCIEX, Concord, ON, Canada) was used for processing (data mining, alignment, normalization, and PCA) of the HPLC–QqQ/ LIT full MS records. Data mining was performed with the use of an automated algorithm in the retention time (RT) range 0.5–7.5 min and m/z range 100–1000. Extraction intensity threshold was set to 10,000 cps, thus only peaks with signal intensity equal or higher than this value were extracted. In the next step, RT and m/z alignment of the respective peaks was carried out using RT and m/z tolerances of 0.2 min and 0.5 Da, respectively. Two separate positive and negative ionization data matrices, comprising lists of peaks characterized for each sample by RT, m/z, intensity and charge state, were obtained and total area sum normalization was performed for each sample. For further statistical evaluation with multivariate methods, positive ionization data, which were found to have higher discriminative potential compared to those recorded in negative ionization mode (as discussed in section below). The data set consisted of total 84 juice samples (orange, $n = 19$; apple, $n = 16$; grapefruit, $n = 16$; orange–apple admixtures, $n = 15$; orange–grapefruit admixtures, $n = 18$). This data set was analyzed with both PCA and LDA. Prior to the actual PCA, the data were pre-processed using the Pareto scaling (the square root of the standard deviation is used as the scaling factor).

Stepwise LDA was performed employing statistiXL (version 1.8, Broadway–Nedlands, Australia) software package; principal components calculated within PCA of the data were used as input variables. Classification results of LDA model were presented in terms of recognition and prediction abilities. Recognition ability represents a percentage of successfully classified samples in the training set. Prediction ability is a percentage of correctly classified samples in the test set by using the model developed during the training step. Validation of the LDA model was carried out following leave-one-out cross-validation (LOOCV) and sevenfold CV procedures. The LOOCV removes only one sample at a time from the training set and considers it as a test set. On the other hand, sevenfold CV consists of assigning samples randomly to a training set and a test set, the latter containing 1/7 of the samples.

To evaluate the high resolution (HR) MS and MS/MS mass spectra obtained within HPLC–QqTOFMS analyses and to estimate elemental formulae of pre-selected marker compounds, PeakView software (version 1.0 with FormulaFinder plug-in version 1.0, AB SCIEX, Concord, ON, Canada) was used.

3 Results and discussion

3.1 HPLC–MS metabolomic analysis

To obtain characteristic metabolomic fingerprints of examined fruit juice samples, high-throughput HPLC–MS method (total run time 10 min), allowing the analysis of a wide range of non-targeted compounds, was developed. For this purpose, generic instrumentation parameter settings were used during analyses for both chromatographic separation and MS detection. Considering the fact that any sample treatment step can potentially result in its alteration and consecutive losses of some metabolites, only few sample preparation steps (i.e., centrifugation, filtration and dilution), essential to avoid instrument problems like column clogging or MS system contamination, were carried out.

Reversed stationary phases (especially C18, but also C8) are widely employed for chromatographic separations in the field of metabolomics, providing acceptable retention of compounds with medium and low polarity. However, these separation systems do not give particularly good results for polar and ionic compounds (Theodoridis et al. 2008). To overcome, at least partly, this drawback, analytical column with stationary phase containing embedded polar groups has been used in this study for the retention increase of polar metabolome components, such as amino acids and sugars. Additionally, weak mobile phase (10% aqueous methanol) was used for elution in initial 3 min of

the chromatographic run. Examples of complex HPLC–QqQ/LITMS total ion current (TIC) chromatograms recorded in positive ionization mode are provided in Fig. 1. Under experimental conditions, the typical base widths of chromatographic peaks were below 20 s. The reproducibility of both m/z and RT plays an important role in successful processing of metabolomic data, especially in the peak aligning step. While the minimization of m/z values fluctuation was ensured by regular mass spectrometer calibration, the in-batch stability of RT, which could be skewed for example by insufficient column equilibration, had to be examined. For this purpose, repeated measurements ($n = 6$) of randomly selected orange juice sample were performed in a batch with all other samples (random sequence) and the RT variability of three peaks eluting in RTs 0.9, 3.5 and 5.2 min was evaluated. Based on these data, typical peak RT variability was found to be below 3 s (standard deviation) or a relative standard deviation less than 3%, which can be considered as acceptable values.

Although the use of HR mass spectrometry utilizing TOF, QqTOF or orbitrap mass analyzers in metabolomic experiments is growing (Werner et al. 2008), in this study, hybrid QqQ/LIT MS system with unit mass resolving power (~ 500–1000 fwhm) was employed for the comprehensive characterization of juices. Besides the lower hardware cost, the main advantage when utilizing mass analyzers operated at low mass resolving power is related to the volume of obtained metabolomic data and consecutively the time needed for its processing, especially during data mining and alignment step, which are in this case significantly lower compared to data acquired with the use of HRMS instruments. It should be noted, however, that HRMS is essential for (tentative) identification of unknown compound of interest, as further discussed and demonstrated in subsection 3.3.

The fast scanning LIT mass analyzer (4000 Da s^{-1}) enabled rapid spectra acquisition and good characterization of analytes peak shape (typically more than 12 points per peak) in full MS mode. Moreover, utilization of the ion trap dynamic fill time function diminished the risk of LIT detector saturation for abundant compounds, and, on the other hand, improved sensitivity for those present at low concentrations. Beside of full scan mass spectra recording, simultaneous automatic acquisition of MS/MS fragmentation mass spectra was performed for ions exceeding the intensity threshold setting, allowing obtaining additional information on respective markers. With regard to the different ionization behavior of various components, juice samples were analyzed both in positive and negative ionization mode, to make the metabolome coverage wider and to assess the suitability of respective data sets for the purpose of fruit juices groups' discrimination.

Fig. 1 HPLC–ESI(+)-QqQ/LITMS total ion current (TIC) chromatograms of fruit juices: **a** orange juice, **b** grapefruit juice, **c** apple juice
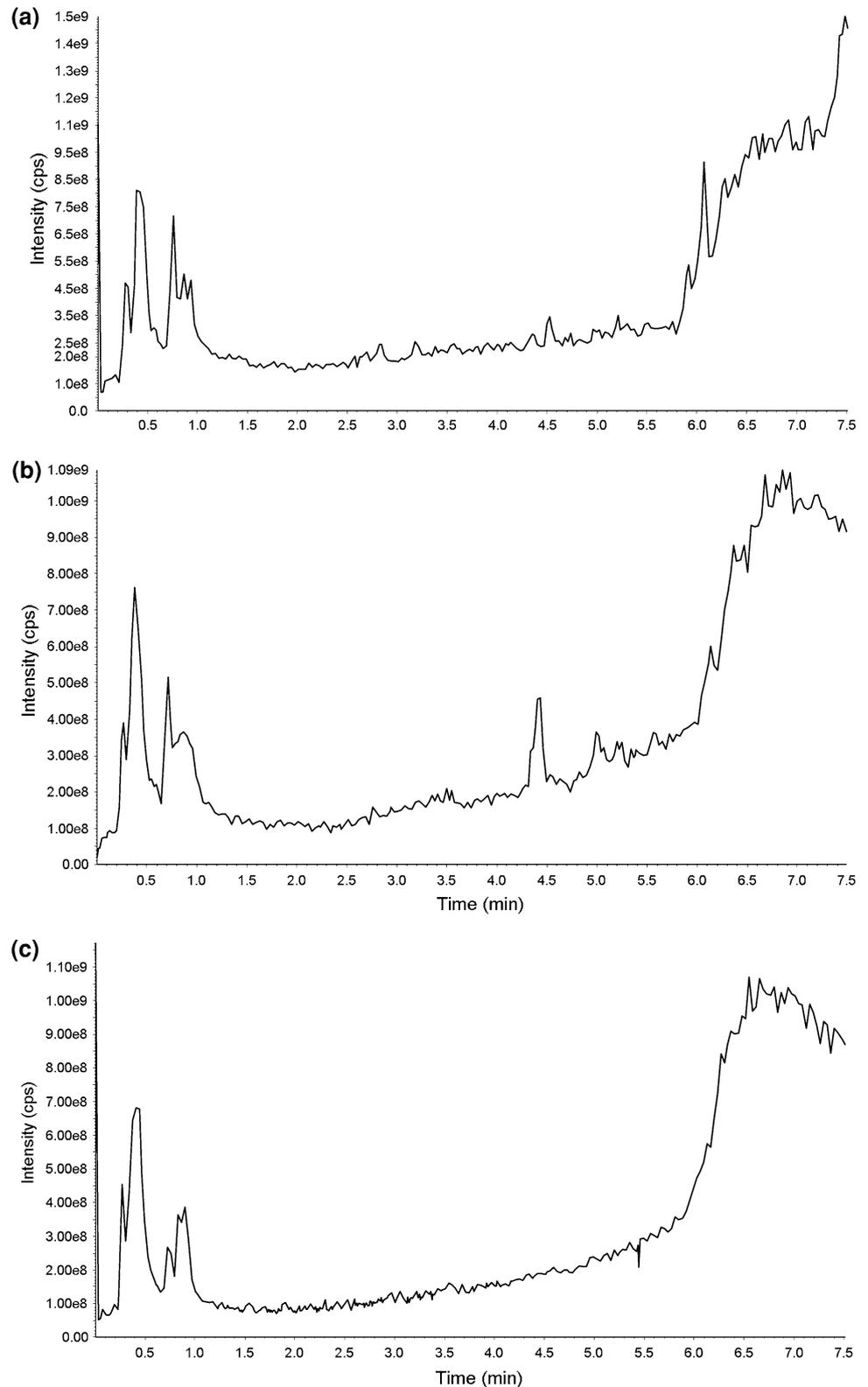
3.2 Chemometric analysis

As an outcome of the data mining and aligning procedures, positive and negative ionization data matrices of 1605 and 4993 peaks, respectively, were obtained. To lower the dimensionality of the data matrices and minimize the signal redundancy, only peaks representing monoisotopic ions (signals with the lowest m/z value within an isotope

pattern), which were assigned during data mining based on isotope spacing, were selected (241 peaks in positive and 1735 peaks in negative ionization data) and subjected to chemometric analysis.

The PCA represents a highly useful tool when interpreting complex multivariate data sets, as it allows dimensionality reduction and visualization of the information present in the original data in the form of a few principal components (PCs) while retaining the maximum possible variability (Berrueta et al. 2007). Large differences in abundance of input variables (peaks) can significantly affect the data variance determined within PCA. To modify the weights of respective variables, Pareto scaling of the data was performed. This type of scaling, which enables reduction, but not complete elimination of input variables abundance differences, provides good results when applied to LC–MS data, as it reflects the fact that larger peaks can be generally considered more reliable, but all variables are equivalent. In this study, PCA was used to get information on the juice samples clustering, as well as for selection of characteristic marker peaks.

As shown in Fig. 2, regardless the ionization mode, PCA of the data set representing pure fruit juices ($n = 51$) revealed three separate clusters of orange ($n = 19$), grapefruit ($n = 16$), and apple juices ($n = 16$) samples, documenting differences in LC–MS profiles associated with the fruit type. It is apparent that more pronounced clustering and significantly better differentiation among sample clusters was obtained for positive ionization data (Fig. 2a) compared to those acquired in the negative mode (Fig. 2c). Therefore, only data recorded in positive ionization mode were further used in this study, as it was found
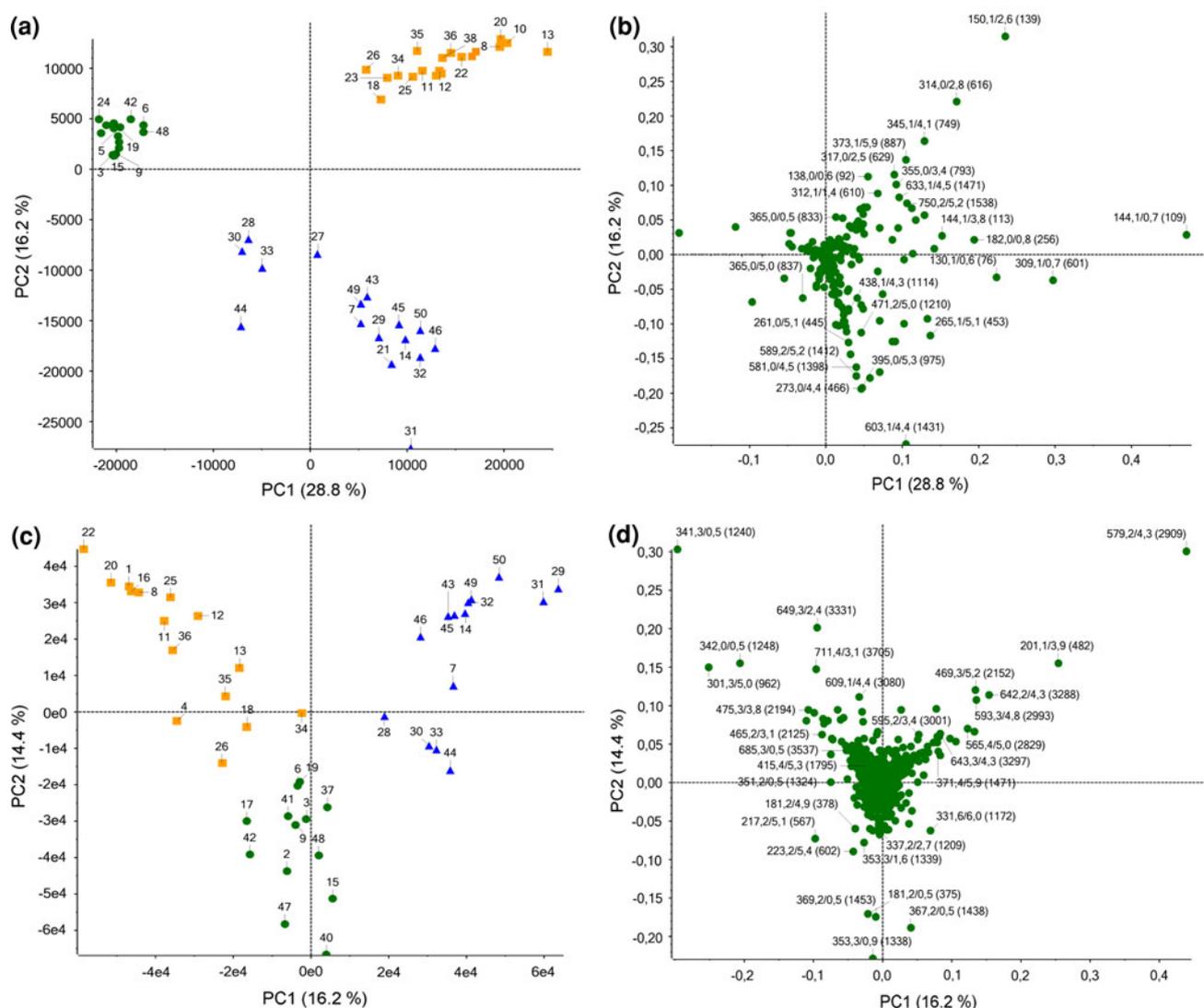
Fig. 2 Results of principal component analysis (*filled square* orange, *filled triangle* grapefruit, *filled circle* apple): **a** Scores plot, positive ionization data, **b** Loadings plot, positive ionization data, **c** Scores plot, negative ionization mode data, **d** Loadings plot, negative ionization data

fit for purpose. It is important to note, however, that comprehensive information in negative electrospray ionization fingerprints can also be useful when dealing with other issues related to fruit juices authenticity and/or during screening for (unknown) adulterants. Within PCA of positive ionization data, 24 PCs explaining 90% of total variance were calculated. In the PCA loadings plot (Fig. 2b), four groups of variables (in total 20 peaks) with high PC1 and/or PC2 absolute loadings values could be observed. Three of these groups were characteristic and responsible for clustering of samples representing respective fruit juice types (i.e., orange, grapefruit, and apple), variables in the fourth group were markers of citrus fruit juices (orange and grapefruit). The overview of marker peaks is provided in Table 2 (see subsection 3.3). The unique occurrence of the markers in tested fruit juices was confirmed by creation of profile plots displaying marker's abundance across the sample set (example in Fig. 3), automatic extraction of corresponding peaks from the raw data, and inspection of blank samples records.

In the final phase of the data analysis, the stepwise forward LDA predictive model for sample classification was developed. Using this supervised pattern recognition method, also the data potential for detection of expensive juice adulteration (orange juice) with cheaper juices (apple and grapefruit) was explored. Beside of three original fruit juice groups (classes), two additional sample classes, representing in laboratory prepared orange–apple and orange–grapefruit admixtures (see Table 1), were employed in LDA. Considering the need for optimal reduction of extremely high number of variables in order to avoid LDA model overfitting, another PCA was performed on the data matrix consisted of 241 signals × 84 samples [containing both pure juice samples ($n = 51$) as well as admixtures ($n = 33$)]. The obtained component scores (PCs) were used as inputs for LDA. The optimal number of component scores was obtained using LOOCV observing a rate of misclassification (from prediction ability). The minimum (optimal) value was obtained for 19 PCs. The misclassification rate was further tested using a sevenfold CV

Table 2 The overview of characteristic marker compounds detected in examined juice types

Marker number	m/z (HRMS)	RT (min)[a]	Ion	m/z (HRMS fragment ions)	Probable ion elemental formula	Mass error (ppm)	Tentative identification	Juice type
1	150.0915	2.6	$[M + H]^+$	135.0666; 107.0478; 91.0522; 77.0366	$C_9H_{12}NO$	+2.0	Unknown	O
2	314.1579	2.8	$[M + H]^+$	152.1060; 121.6035	$C_{15}H_{24}NO_6$	−2.6	Unknown	O
3	317.1180	2.5	$[M + H]^+$	203.40496	$C_{21}H_{17}O_3$	+2.5	Unknown	O
4	345.1510	4.1	$[M + Na]^+$	n.d.	$C_{14}H_{26}O_8Na$	−2.9	Unknown	O
5	373.1276	5.9	$[M + H]^+$	358.1021; 343.0787	$C_{20}H_{21}O_7$	−1.6	Tangeretin	O
6	633.1794	4.5	$[M + Na]^+$	611.1962; 449.1433; 303.0856	$C_{28}H_{34}O_{15}Na$	+0.6	Hesperidin	O
7	261.1119	5.1	Unknown	189.0548; 131.0484	$C_{15}H_{17}O_4$	−0.8	Unknown	G
8	273.0750	4.4	$[M + H]^+$	153.0172; 147.0417; 119.0468	$C_{15}H_{13}O_5$	−2.7	Naringenin	G
9	301.1033	5.1	$[M + H]^+$	240.2319; 213.0057	$C_{12}H_{17}N_2O_7$	−1.8	Unknown	G
10	471.2024	5.0	$[M + H]^+$	425.1975; 161.0586	$C_{26}H_{31}O_8$	+2.3	Limonin	G
11	581.1854	4.4	$[M + H]^+$	273.0742	$C_{27}H_{33}O_{14}$	−1.9	Naringin	G
12	589.3281	5.2	$[M + Na]^+$	571.3191	$C_{31}H_{50}O_9Na$	−1.2	Unknown	G
13	603.1696	4.4	$[M + Na]^+$	n.d.	$C_{27}H_{32}O_{14}Na$	+1.9	Naringin	G
14	130.0859	0.6	$[M + H]^+$	84.0806; 82.0657	$C_6H_{12}NO_2$	−3.1	N-methyl proline	O, G
15	144.1011	0.7	M^+	84.0449; 58.0645	$C_7H_{14}NO_2$	−5.6	Proline betaine	O, G
16	182.0821	0.8	$[M + H]^+$	136.0749	$C_9H_{12}NO_3$	+4.9	Tyrosine	O, G
17	449.1760	3.9	$[M + Na]^+$	287.1238	$C_{21}H_{30}O_9Na$	−4.9	Unknown	O, G
18	203.0522	0.5	$[M + Na]^+$	n.d.	$C_6H_{12}O_6Na$	−1.8	Hexose	A
19	205.0675	0.5	$[M + Na]^+$	n.d.	$C_6H_{14}O_6Na$	−3.7	Sorbitol/mannitol/ dulcitol	A
20	541.2248	3.5	$[M + H]^+$	409.1816	$C_{22}H_{33}N_6O_{10}$	−0.9	Unknown	A

O orange, G grapefruit, A apple, n.d. not detected

[a] Retention time

Fig. 3 Responses of selected markers in fruit juices (*filled square* orange, *filled triangle* grapefruit, *filled circle* apple): **a** *m/z* 314.0, RT 2.8 min; **b** *m/z* 273.0, RT 4.4 min; **c** *m/z* 130.1, RT 0.5 min; **d** *m/z* 205.0, RT 0.5 min
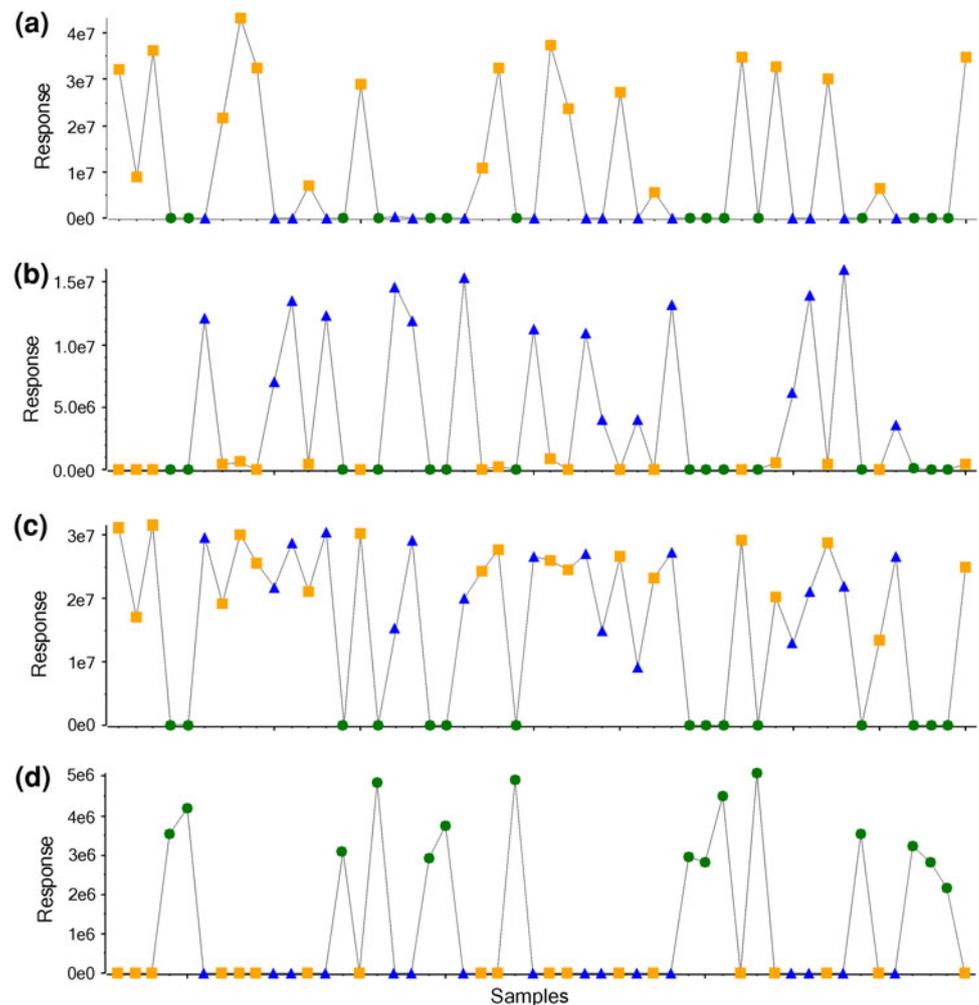
procedure. In both cases, the sample(s) from the validation set were classified using LDA on the bases of Bayesian decision theory [i.e., membership of each class has a prior probability (in particular $P = 0.2$)]. In addition, the possibility of LDA model overfitting was also checked by using the criterion $(n - g)/3$ (n is the number of objects and g represents the number of classes), which should not be lower than the number of variables used. While the recognition ability of LDA model constructed with the use of the whole data set was 100%, the prediction ability calculated during LOOCV and sevenfold CV was lower (94.1 and 92.9%, respectively). In the case of LOOCV, 3 orange–apple (90:10, *v/v*) mixtures and 2 orange–grapefruit (90:10, *v/v*) mixtures were misclassified. In the case of sevenfold CV, misclassification of all orange–apple and orange–grapefruit mixtures prepared at the ratio 90:10 (*v/v*) was observed. Correct classification (recognition and prediction abilities of 100% for both LOOCV and sevenfold CV) of all samples in both training and validation step was achieved when orange–apple/orange–grapefruit admixtures prepared at adulteration level 10% were removed from the

data. The final LDA model enabled reliable detection of 15% addition of apple and grapefruit juice into orange juice, the presence of orange juice in grapefruit and apple juice was detectable at 25 and 10%, respectively. Figure 4 shows a scores plot of three discriminant functions documenting satisfactory resolution among the classes.

3.3 Tentative identification of marker compounds

The identification of discriminating marker compounds represents probably the most laborious and time-consuming step of the metabolomic workflow. As mentioned above, the use of a high-resolution instrument enabling to obtain both single MS and MS/MS accurate mass spectra is needed for reliable elemental formula estimation, which is typically followed by a database search. Of course, the HPLC–QqQ/LITMS data collected during the first phase of this study did not provide enough information to accomplish this task; therefore, further analyses of selected fruit juice samples (each one representing orange, grapefruit, or apple juice) were performed using an HPLC–QqTOFMS
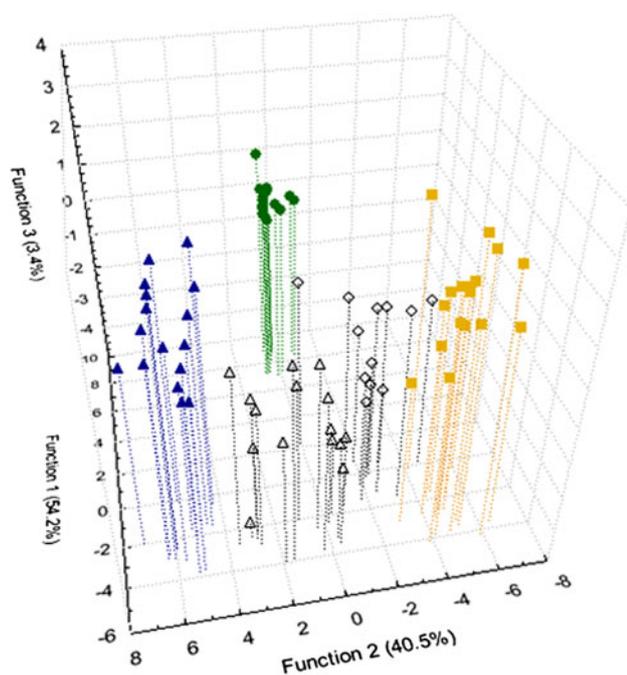
Fig. 4 3D scores plot of three discrimination functions (*filled square* orange, *filled triangle* grapefruit, *filled circle* apple, *open diamond* orange–apple mixes, *open triangle* orange–grapefruit mixes)

mass spectrometer. It should be noted that the confirmatory analysis of a predicted compound standard, which is generally required for full and unequivocal identity confirmation, was not carried out in this study. With regard to this fact, the identification performed here can be considered only tentative.

Within the initial stage of the HPLC–QqTOFMS records examination, a search for marker compounds peaks was carried out. Using the information on RT and *m/z* values obtained from HPLC–QqQ/LITMS data, all the peaks of interest could be successfully found in HRMS chromatograms at almost identical RTs. To gain higher level of confidence and to avoid false positive peak assignment, also MS/MS fragmentation spectra, automatically acquired by both instruments, were compared. Although the respective fragmentation patterns differed slightly in terms of fragment ions relative intensities, generally a good match between low resolution (LR) and HR MS/MS mass spectra was observed and confirmation of markers was possible. It is well known, that besides protonated molecules $[M + H]^+$, also various adduct ions, such as $[M + Na]^+$, $[M + K]^+$, and/or $[M + NH_4]^+$, can be formed from a single molecule, when positive polarity electrospray ionization technique is employed. The correct assignment of an ion type is an important step in unknowns' identification, since it can significantly reduce the number of candidates provided by accurate mass-based elemental formula estimation. With regard to this fact, ion

type assignment was performed for each marker's monoisotopic signal based on a search for *m/z* values corresponding to above listed adducts, exhibiting the same RT and elution profile (accurate *m/z* differences of ± 21.9819, ± 37.9559, ± 17.0265, ± 4.9554, ± 20.9294 from observed value were considered). As documented in Table 2, using this approach, ion type could be assigned to most of the marker compounds, thanks to the presence of related protonated molecules and/or adduct ions.

The automatic elemental formula estimation was carried out with the use of: (i) accurate single MS mass of parent ion, (ii) isotopic profile of parent ion, and (iii) accurate MS/MS fragment ions. Unfortunately, as expected, under experimental condition some of $[M + Na]^+$ adducts did not provide any fragment ions, thus only parent ion could be used to predict their elemental composition. For calculations, C ($n \leq 50$), H ($n \leq 100$), N ($n \leq 10$), O ($n \leq 20$), P ($n \leq 15$), and S ($n \leq 5$) atoms were considered; during formula estimation of ions previously assigned as sodium adducts, Na atom ($n = 1$) was added to the above list. The FormulaFinder software used in this study enabled to sort proposed formulas according to "MS rank" and "MS/MS rank" (the lower the rank, the higher the probability that proposed formula is correct), reflecting the differences between calculated and measured *m/z* values for both parent and fragment ion(s), and the match of experimental and theoretical isotope pattern in terms of isotope spacing and relative intensities. Due to rather narrow quadrupole isolation window (0.7 Da) setting, which was used in order to obtain clean fragmentation spectra, the isotope pattern information could be utilized only for parent ions. In the next step, stepwise search of the markers' molecular formulae (derived from estimated elemental formulae by subtracting hydrogen or sodium atom) in several online databases [MassBank (http://www.massbank.jp/?lang=en) Metlin (http://metlin.scripps.edu/metabo_search_alt2.php), and PubChem (http://pubchem.ncbi.nlm.nih.gov/)] was performed in the ascending order according to MS (and MS/MS) ranks. Of the numerous compound hits obtained, only candidates, presence of which was probable in fruit juices, were further examined by comparing experimental fragmentation mass spectra with those provided either in databases (MassBank and Metlin) and/or scientific literature. Using this procedure, 11 of 20 marker compounds were tentatively identified (see Table 2).

In Fig. 5, accurate MS/MS mass spectra examples of four selected markers are shown. The search for molecular formula, which was obtained based on mass spectrum shown in Fig. 5a, provided only 1 hit in MassBank database. A comparison of reference and experimental MS/MS spectra have confirmed that the unknown compound was probably naringenin, a flavanone found at high concentrations in grapefruits. Similar mass spectra, dominated by
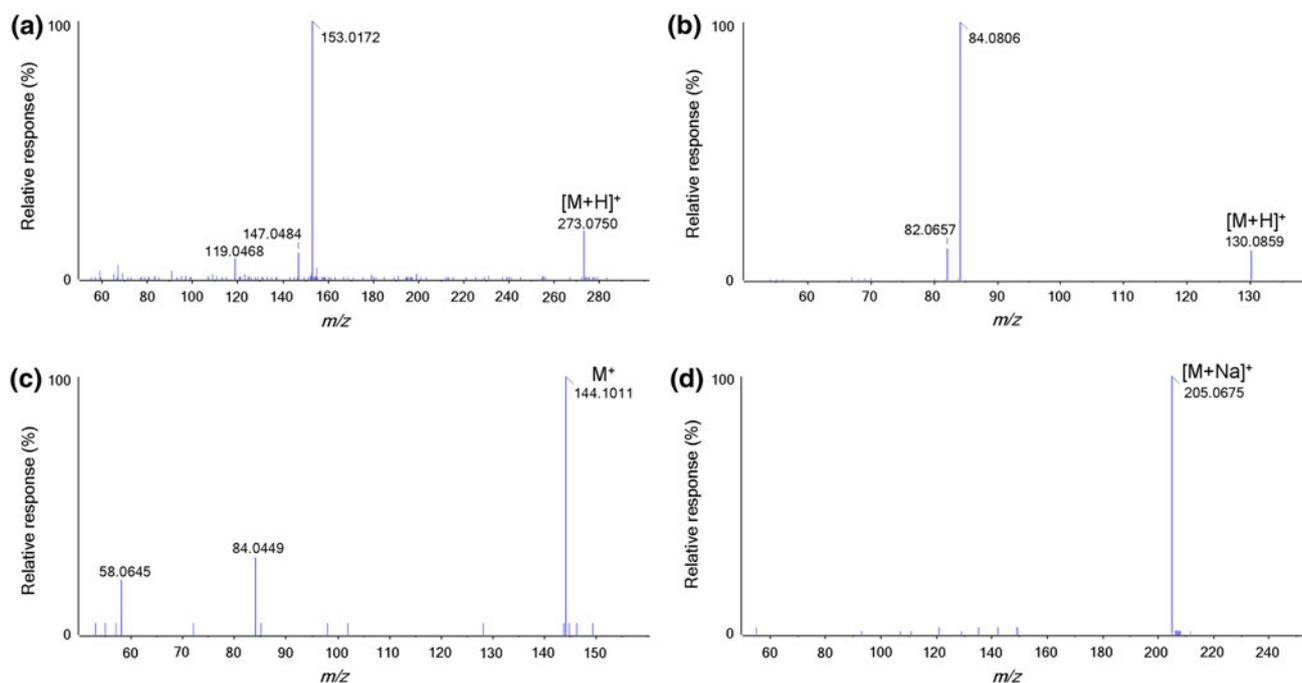
Fig. 5 Examples of ESI(+) HR MS/MS mass spectra of selected markers tentatively identified in fruit juices: **a** naringenin, **b** N-methyl proline, **c** proline betaine, **d** sorbitol/mannitol/dulcitol

$[M + H - C_8H_9O]^+$ ion (m/z 153.0172) were also found in the literature (Abad-García et al. 2009). Using the accurate MS/MS spectra characterized by two fragment ions at m/z 84.0806 and 82.0657 (Fig. 5b), the best ranked elemental composition $C_6H_{12}NO_2$, was estimated. Corresponding molecular formula gave multiple hits in all of searched databases, while no available reference MS/MS spectra matched the experimental fragmentation pattern. However, with exception of one compound (N-methyl proline), all of these substances were either unlikely to be abundant in citrus juices or could not yield observed fragment ions pattern. Considering the structure of N-methyl proline, one of candidates provided by the Pub-Chem database, decarboxylation and double bond formation from parent $[M + H]^+$ ion could be attributed to respective fragments. The identification of this marker as N-methyl proline is also supported by almost identical MS/MS spectral data published by Servillo et al. (2011). In case of spectra displayed in Fig. 5c, the best ranked elemental formula $C_7H_{14}NO_2$ was obtained for M^+ ion. The Metlin database provided 2 hits. While the reference MS/MS spectra, which were available for only one of the candidates, did not match the experimental data, the second candidate, proline betaine, was previously shown to yield observed fragment ions through a rearrangement fragmentation process resulting in the loss of C_3H_6 followed by loss of water (Wood et al. 2002). As can be seen in Fig. 5d, no fragment ions were formed from the $[M + Na]^+$ adduct

ion. Molecular formula $C_6H_{14}O_6$ (derived from second best ranked elemental formula) was used to search the databases and three identical candidates, representing isomeric sugar alcohols (sorbitol, mannitol, and dulcitol), were found. However, using the available experimental data, no further candidates could be excluded.

4 Conclusions

Comprehensive, non-targeted HPLC–MS metabolomic fingerprinting coupled to chemometric methods was demonstrated to be a powerful tool for fruit juices authenticity testing. With the use of LR MS instrument utilizing QqQ/LIT mass analyzer, information-rich, discriminating records were collected for variable sample set of orange, grapefruit and apple juices and subsequently subjected to PCA and LDA. Based on LDA classification model constructed with the use of positive electrospray ionization data, reliable detection of orange juice adulteration with apple and grapefruit juice was possible at 15% addition level. Finally, an attempt to tentatively identify marker compounds, preselected during PCA, was carried out based on single MS and MS/MS mass spectra obtained within the HPLC–QqTOFMS analysis. Using a combination of experimental data and information available in chemical, metabolomic and mass spectral databases, more than 50% of characteristic markers were tentatively identified.

References

Abad-García, B., Berrueta, L. A., Garmón-Lobato, S., Gallo, B., & Vicente, F. (2009). A general analytical strategy for the characterization of phenolic compounds in fruit juices by high-performance liquid chromatography with diode array detection coupled to electrospray ionization and triple quadrupole mass spectrometry. *Journal of Chromatography A, 1216*, 5398–5415.

Berrueta, L. A., Alfonso-Salces, R. M., & Hérberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*, 196–214.

Catillo, M. L. R., Caja, M. M., & Heraiz, M. (2003). Use of enantiomeric composition for the assessment of the authenticity of fruit beverages. *Journal of Agricultural and Food Chemistry, 51*, 1284–1288.

Cevallos–Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., & Rodrick, G. E. (2009). Metabolomic analysis in food science: A review. *Trends in Food Science & Technology, 20*, 557–566.

Council Directive 2001/112/EC relating to fruit juices and certain similar products intended to human consumption. *Official Journal of the European Communities*, L10/58–L10/66.

Cuny, M., Vigneau, E., Le Gall, G., Colquhoun, I., Lees, M., & Rutledge, D. N. (2008). Fruit juice authentication by ^{1}H NMR spectroscopy in combination with different chemometric tools. *Analytical and Bioanalytical Chemistry, 390*, 419–427.

Dunn, W. B., & Ellis, D. I. (2005). Metabolomics: Current analytical platforms and methodologies. *TrAC-Trends in Analytical Chemistry, 24*, 285–294.

Ebeler, S. E., Takeoka, G. R., & Winterhalter, P. (2007). *Authentication of food and wine*. Washington DC: American Chemical Society.

Ehling, S., & Cole, S. (2011). Analysis of organic acids in fruit juices by liquid chromatography-mass spectrometry: An enhanced tool for authenticity testing. *Journal of Agricultural and Food Chemistry, 59*, 2229–2234.

Fügel, R., Carle, R., & Schreiber, A. (2005). Quality and authenticity control of fruit purées, fruit preparations and jams—A review. *Trends in Food Science & Technology, 16*, 433–441.

Gómez-Ariza, J. L., Villegas-Portero, M. J., & Bernal-Daza, V. (2005). Characterization and analysis of amino acids in orange juice by HPLC-MS/MS for authenticity assessment. *Analytica Chimica Acta, 540*, 221–230.

He, J., Rodriguez-Saona, L. E., & Giusti, M. M. (2007). Midinfrared spectroscopy for juice authentication-rapid differentiation of commercial juices. *Journal of Agricultural and Food Chemistry, 55*, 4443–4452.

Jezek, J., & Suhaj, M. (2001). Application of capillary isotachophoresis for fruit juice authentication. *Journal of Chromatography A, 916*, 185–189.

Kaddurah-Daouk, R., & Krishnan, K. R. (2008). Metabolomics: A global approach to the study of central nervous system diseases. *Neuropsychopharmacology, 34*, 173–186.

Kelly, J. F. D., & Downey, G. (2005). Detection of sugar adulterants in apple juice using Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 53*, 3281–3286.

Le Gall, G., Puaud, M., & Colquhoun, I. J. (2001). Discrimination between orange juice and pulp wash by ^{1}H nuclear magnetic resonance spectroscopy: Identification of marker compounds. *Journal of Agricultural and Food Chemistry, 49*, 580–588.

Low, N. H. (1996). Determination of fruit juice authenticity by capillary gas chromatography with flame ionization detection. *Journal of AOAC International, 79*, 724–737.

Muntean, E. (2010). Simultaneous carbohydrate chromatography and unsuppressed ion chromatography in detecting fruit juices adulteration. *Chromatographia, 71*, S69–S74.

Pan, G. G., Kilmartin, P. A., Smith, B. G., & Melton, L. D. (2002). Detection of orange juice adulteration by tangelo juice using multivariate analysis of polymethoxylated flavones and carotenoids. *Journal of the Science of Food and Agriculture, 82*, 421–427.

Saavedra, L., García, A., & Barbas, C. (2000). Development and validation of a capillary electrophoresis method for direct measurement of isocitric, citric, tartaric and malic acids as adulteration markers in orange juice. *Journal of Chromatography A, 881*, 395–401.

Saavedra, L., Rupérez, F. J., & Barbas, C. (2001). Capillary electrophoresis for evaluating orange juice authenticity: A study on Spanish oranges. *Journal of Agricultural and Food Chemistry, 49*, 9–13.

Servillo, L., Giovane, A., Balestrieri, M. L., Cautela, D., & Castaldo, D. (2011). Proline derivatives in fruits of bergamot (*Citrus bergamia* Risso et Poit): Presence of *N*-methyl-L-proline and 4-hydroxy-L-prolinebetaine. *Journal of Agricultural and Food Chemistry, 59*, 274–281.

Shaw, P. E., Moshonas, M. G., Buslig, B. S., Barros, S. M., & Widmer, W. W. (1999). Discriminant and principal component analysis to classify orange juices based on relative volatile juice constituents. *Journal of the Science of Food and Agriculture, 79*, 1949–1953.

Theodoridis, G., Gika, H. G., & Wilson, I. D. (2008). LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *TrAC Trends in Analytical Chemistry, 27*, 251–260.

Tønder, D., Petersen, M. A., Poll, L., & Olsen, C. E. (1998). Discrimination between freshly made and stored reconstituted orange juice using GC odour profiling and aroma values. *Food Chemistry, 61*, 223–229.

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.

Wishart, D. S. (2008). Application of metabolomics in drug discovery and development. *Drugs in R&D, 9*, 307–322.

Wood, K. V., Bonham, C. C., Miles, D., Rothwell, A. P., Peel, G., Wood, B. C., et al. (2002). Characterization of betaines using electrospray MS/MS. *Phytochemistry, 59*, 759–765.