COST STSM: Preparation of a manuscript based on the results from the multiplatform analysis of plant bioactives

Scientific report

1 Location and duration of work

- iBET/ITQB NOVA, Oeiras, Portugal
- December 4 to 14, 2017

2 Objective

The aim of this short-term scientific mission was to prepare the first draft of a manuscript initially titled *“European multiplatform strategy for a coverage test on plant bioactives and their metabolites by mass spectrometry”*, which is based on the results from the multiplatform analysis of phytochemical standards analysed in 10 different platforms during 2017. This work is related to the previous STSM by the same applicant carried out in the same location in October 2016 for the preparation of the standard mixtures. The manuscript is planned to be published in a top field-specific peer-reviewed journal, possibly as an open-access article, during early 2018.

3 Materials and methods

Prior to the STSM (and during it for double-checking), the results data sent by the platforms were processed and formatted into one Excel file. The processing included the determination of positive identification and the collection of relevant identification data, such as retention times, exact measured *m/z*, MS/MS fragmentation, and S/N ratios. The rough content of the manuscript was also agreed upon before the start of the STSM. During the trip, two online meetings were held on December 11 and 14 (participants: Maria Rosário Bronze, Ville Koistinen, Dorrain Low-Yanwen, Claudine Manach) to review the current progress and decide on the directions for continuing the writing work.

4 Results and discussion

A literature review was performed on the targeted and untargeted analysis of various groups of phytochemicals to include in the background section of the manuscript. Information regarding the platform specifications were collected into a table. To show the main identification results, another table was created in Excel, and based on the comments from co-authors, it was transformed into a colour-coded image for more clear visual interpretation (Figure 2 in the manuscript). The chemical space image (Figure 1) originally prepared by Andreia Bento da Silva during her STSM was updated and modified to include both the

analysed standards and the original list of 162 compounds. Additional figures were created for the chromatographic and MS separation of certain compounds as well as a graphic to visualize the different gradients used by the LC–MS platforms. In the results and discussion, the following aspects of the results were discussed: the identification of compounds in different chemical classes, instruments, ionization modes and mixtures, the reliability of the identifications, the influence of chromatography and the ionization mode, and considerations on the optimal coverage of plant bioactives for untargeted analyses, including a suggested set of parameters and standards to be used in all platforms.

The first draft of the manuscript, prepared during the STSM, is included in the Appendix. Some parts of the manuscript still require additions / shortening based on Claudine's comments, after which the manuscript will be sent for limited circulation to the main authors and then to all the co-authors mentioned in the paper. The deadline for the submission of the paper is January 2018.

5 Appendix: the manuscript draft

European multiplatform strategy for a coverage test on plant bioactives and their metabolites by mass spectrometry

Ville Mikael Koistinen[1], Andreia Bento da Silva[2], László Abrankó[3], Rocio Garcia Villalba[4], Francisco Tomás Barberán[4], Rikard Landberg[5], Christof Van Poucke[6], Kateřina Valentová[7], Stephanie Durand[8], Dorrain Low-Yanwen[8], Rafael Llorach[9], Mar Garcia-Aloy[9], Mireia Urpi-Sarda[9], Anna-Marja Aura[10], Tuulikki Seppänen-Laakso[10], Kati Hanhineva[1], Claudine Manach[8], Maria Rosário Bronze[2,11,12]

1 University of Eastern Finland, Institute of Public Health and Clinical Nutrition, P.O. Box 1627, FI-70211 Kuopio, Finland

2 Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa (ITQB NOVA), Av. da República, 2780-157 Oeiras, Portugal

3 Szent István Egyetem, Elelmiszertudomanyi Kar, Alkalmazott Kemia Tsz., 29-43 Villanyi street, 1118 Budapest, Hungary

4 CEBAS-CSIC, Campus Universitario de Espinardo, edf 25, 30100 Murcia, Spain

5 Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

6 Institute for Agricultural and Fisheries Research (ILVO), Technology and Food Science Department, Brusselsesteenweg 370, B-9090 Melle, Belgium

7 Laboratory of Biotransformation, Vídeňská 1083, CZ-142 20 Prague, Czechia

8 INRA, Human Nutrition Unit, Université Clermont Auvergne, F63000 Clermont-Ferrand, France

9 University of Barcelona, Nutrition, Food Science and Gastronomy Department, Pharmacy Faculty, University of Barcelona. Av Joan XXIII, s/n, 08028 Barcelona, Spain

10 VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Espoo, Finland

11 Instituto de Biologia Experimental Tecnológica (iBET), Apartado 12, 2781-901, Oeiras, Portugal

12 Faculty of Pharmacy, University of Lisbon (FFULisboa), Avenida Professor Gama Pinto, 1649-003 Lisbon, Portugal

ABSTRACT

State-of-the-art analytical methodologies, such as mass spectrometry and NMR, are required for the profiling of compounds in complex matrices, including plant food materials and biofluids. Especially in untargeted LC–MS, there is need for an optimized method for the detection and identification of hundreds known plant food bioactives and their metabolites. The aim of this project was to compare the analytical coverage of routine untargeted methods developed independently in various European platforms for use in metabolomics studies. Briefly, 56 chemical standards, representing the most common classes of phytochemicals, were dissolved into appropriate solvents and urine, and the standard mixtures were analyzed in the participating platforms ($n = 10$) using their preferred method. The results will serve as a basis for the implementation of a consensus method and definition of criteria to assess the quality of the analytical coverage.

BACKGROUND

Points to address: (i) the importance of mass spectrometry and other methodologies in the identification of compounds; (ii) equipments used, differences and type of response we can get; (iii) does the sample matrix interfere in the analysis? (iii) why we decided it was important to do a study like this?

Plant bioactives, or phytochemicals, are plant-synthetized chemical compounds that do not act as nutrients but may have other biological activity when introduced to the human body within food. The analysis of phytochemicals from food and biofluids provides information about the differences between food varieties, the impact of food processing on the product, and linking the exposure to certain foods with health outcomes. Mass spectrometry (MS) has become an indispensable technology in analysing plant bioactives, or phytochemicals, from various matrices [1-6]. In particular, liquid chromatography coupled with mass spectrometry (LC–MS) provides high speed, mass accuracy, dynamic range and sensitivity due to advances in the instrumentation during the past decade, such as the development of state-of-the-art mass analyser techniques quadrupole time-of-flight (qTOF), Orbitrap and ion cyclotron resonance (ICR). These characteristics are especially valuable in untargeted plant metabolomics, where each sample typically contains a high number of metabolites [7]. Gas chromatography (GC–MS) has a clear advantage over LC–MS in being highly reproducible regarding retention times and the mass spectral fingerprints of compounds [8] and GC coupled with time-of-flight mass spectrometry (GC–TOF–MS) emerged as the first method for a large-scale analysis of plant metabolites [7]; however, the requirement of derivatization to increase the volatility of the analytes limits the usability of the technology to mainly primary metabolites [6, 7]. Diode array detection (DAD), with or without coupling with mass spectrometry, has some useful applications in the analysis of plant bioactives [9, 10]. Mass

spectrometry provides limited structural information on molecules and relies on comparison with mass spectra of previously identified compounds; therefore, complementary methods, such as nuclear magnetic resonance (NMR) spectroscopy, may be used for the complete structural elucidation.

Several targeted methods have been developed for different phytochemical families. Flavonoids have been studied most extensively with different LC–MS techniques [6, 12, 13]. For LC–MS, the clear majority of analyses have been carried out using a reversed-phase (RP) C18 column, but more recently, hydrophilic interaction chromatography (HILIC) has emerged as an option for the analysis of glycosylated flavonoids to complement the results from the traditional RP chromatography [6]. Due to the low volatility of TMS-derivatized flavonoid glycosides, GC–MS is not often used in the analysis of flavonoids [12], unless only the aglycones are among the compounds of interest. Several MS-based methods exist for the analysis of phenolic acids, including UHPLC–MS/MS [14], HPLC–DAD–MS [15, 16], and capillary gas chromatography combined with mass spectrometry (CGC–MS) [17]. In addition, capillary electrophoresis (CE) has been used as the separation method for phenolic acids and UV (or UV-vis) absorption or fluorescence as the detection method [18, 19]. In cereal grains, phenolic acids are mostly bound to hemicelluloses or amines [19], complicating the sample preparation process. Reversed-phase LC–MS is the prevailing method in the analysis of lignans from plant extracts, foods and biological fluids; GC–MS, HPLC–UV and HPLC–DAD are also highly usable [20]. A successful targeted analysis of lignans requires multi-step extraction and sample preparation processes , which may pose challenges for an untargeted study setting, where a simple extraction procedure is usually preferred. GC–MS used to be the most frequently utilized technique for the detection of plant sterols [21]. As with several other groups of plant bioactives, the advances in LC–MS have increased its utilization in the analysis of these compounds [22]. Phase II metabolites of plant bioactives, such as glucuronides and sulfates, have been analyzed with LC–MS/MS from e.g. polyphenols from cocoa [23] and tea [24].

Alkylresorcinols are phenolic lipids that among edible plants exist almost exclusively in the outer layers of wheat, rye and barley grains and have therefore been used as biomarkers of whole-grain intake [25]. The current methods for the specific analysis of alkylresorcinols include GC–MS and normal-phase LC–MS, preceded by sample preparation involving a triplicate liquid–liquid extraction. Non-targeted metabolomics approaches applying LC–MS with reversed-phase chromatography have also detected these compounds from plant matrices [26]. Betaines are hydrophilic compounds containing a quaternary ammonium group and acting as osmotic regulators and methyl donors in mammals. Different combinations of HPLC, CE, UV and MS have been applied in their analysis [27]. Although it is possible to

separate betaines with a reversed-phase column, MS coupled with HILIC chromatography provides superior separation and sensitivity due to the high hydrophilicity of the compounds [28].

The goal of this paper is to describe the procedure used to develop a consensus method (or more likely a combination of methods) with a wide coverage of plant food bioactive metabolites. The procedure would be used e.g. to measure the actual exposure of individuals to plant food bioactives metabolites in biofluids. The test will allow to better assess the analytical coverage (qualitative analysis) of the various instruments and methods used (including GC–MS, GC×GC–TOF, LC–QqQ, LC–qTOF, LC–Orbitrap, and NMR methods). This test, complemented by a literature search to compile knowledge on analytical features of phytochemical and their metabolites, will serve as a basis to establish a consensus multi-platform method for a maximal coverage of phytochemicals and their metabolites in untargeted metabolomics studies.

MATERIALS AND METHODS

Preparation of the test and planning the strategy

The criteria for selecting the chemical standards for the analysis were (i) to represent the consensus interests of the platforms involved, (ii) to represent a variety of food products; (iii) to include compounds from different chemical classes with different chemical characteristics; (iv) to be relatively inexpensive and easy to obtain; (v) to be stable at room temperature. An initial list of 179 compounds covering a wide range of masses and polarities was established based on research interests and previous knowledge of plant bioactives (Supplementary Table 1). A questionnaire was sent to each participating metabolomics platform (n = 18) to determine the availability of these chemical standards and any other standards of plant bioactives that the platforms may have. Based on the answers from the platforms (n = 14), the list was narrowed down to 56 representative compound standards (Supplementary Table 1), which were sent to the organizing laboratory for the preparation of the analytical mixtures. Along with the mixtures, a SOP (Supplementary Figure 1) was sent electronically to the platforms, describing all the procedures carried out by the organizing and participating laboratories as well as instructions and templates for providing the results in a uniform manner.

Reference standard mixtures

56 chemical standards representing the most common phytochemical classes were acquired (1 to 2 mg each) from the participating platforms, originating mostly from commercial vendors and 9 compounds having been synthesised in-house (Supplementary Table 1). The log P values were calculated for each compound using ALOGPS (Figure 1).

Based on the log *P* value and literature and database searches, the solubility of the compounds in common solvents were assessed. Consequently, the standards were dissolved into the most appropriate solvent (water:acetonitrile 1:1, methanol, DMSO, and chloroform) and further classified into three groups/mixtures according to their optimal solubility in solvent A (water:acetonitrile 1:1), B (methanol), or C (chloroform). The target concentration for the stock solutions was 10 mM; in case of incomplete dissolution or precipitation after visual inspection, the suspension was diluted into 5 mM with a less polar solvent or into 0.5 or 1 mM with the same solvent, again based on previous knowledge. Ellagic acid was dissolved in methanol into a concentration of 50 ppm, as this was the maximum concentration possible to achieve based on previous knowledge. Three standard mixtures (A, B and C) were prepared from the stock solutions at 200 µM (except for the less soluble compounds: apigenin 7-*O*-glucoside, bergaptol, catechol-*O*-sulfate, curcumin, ellagic acid, isorhamnetin, luteolin, 5-pentacosylresorcinol, quercetin 3'-*O*-sulfate, quercetin 4'-*O*-sulfate, quercetin disulfate, and sinapic acid). The mixtures were sent to all participating platforms with separate rat urine samples for the analysis of the standards in solvent and urine.

Sample preparation and analysis with mass spectrometry

Each participating platform used their own traditional analytical procedure for analysing the reference standard mixtures and identifying the analytes. The types of equipment used and the conditions for the chromatographic separation and mass spectrometry are listed in Table 1. Before the analysis, the mixtures were diluted into a final concentration of 10 mM using water (mixture A) or methanol (mixtures B and C) as a solvent. Each LC–MS platform was asked to report the following results for each identified compound: retention time, ion mode, observed *m/z* with maximum accuracy, mass error, collision induced dissociation (CID) energy, MS/MS fragments, and signal-to-noise ratio. For the MS/MS data, 10 most intense fragments (or as many observed) with their relative intensity in percentage was to be reported. In the case of discrepancy between the identification and the reported *m/z*, retention time or MS/MS fragmentation pattern, the platform was contacted for verification.

RESULTS AND DISCUSSION

Chemical space. The chemical standards selected for the analysis, including the representative compounds in the initial listing, were plotted based on their molecular mass and calculated log *P* value to visualize the chemical space of the phytochemical families (Figure 1). As indicated by the area to which compounds within the same chemical group have spread, flavonoids are a chemically diverse group, ranging from relatively lipophilic aglycones, such as xanthohumol, to more complex glucosides (naringin). Correspondingly, the addition of a hydrophilic quinic acid moiety in phenolic acids, such as chlorogenic acid,

increases the hydrophilicity despite of the increased mass. Sulfated and glucuronidated phytochemicals are spread over a wide mass range but all have a relatively low log *P* value. Certain lipophilic families, including alkylresorcinols, carotenoids and phytosterols, and hydrophilic betaines, form very distinct groups in the plot.

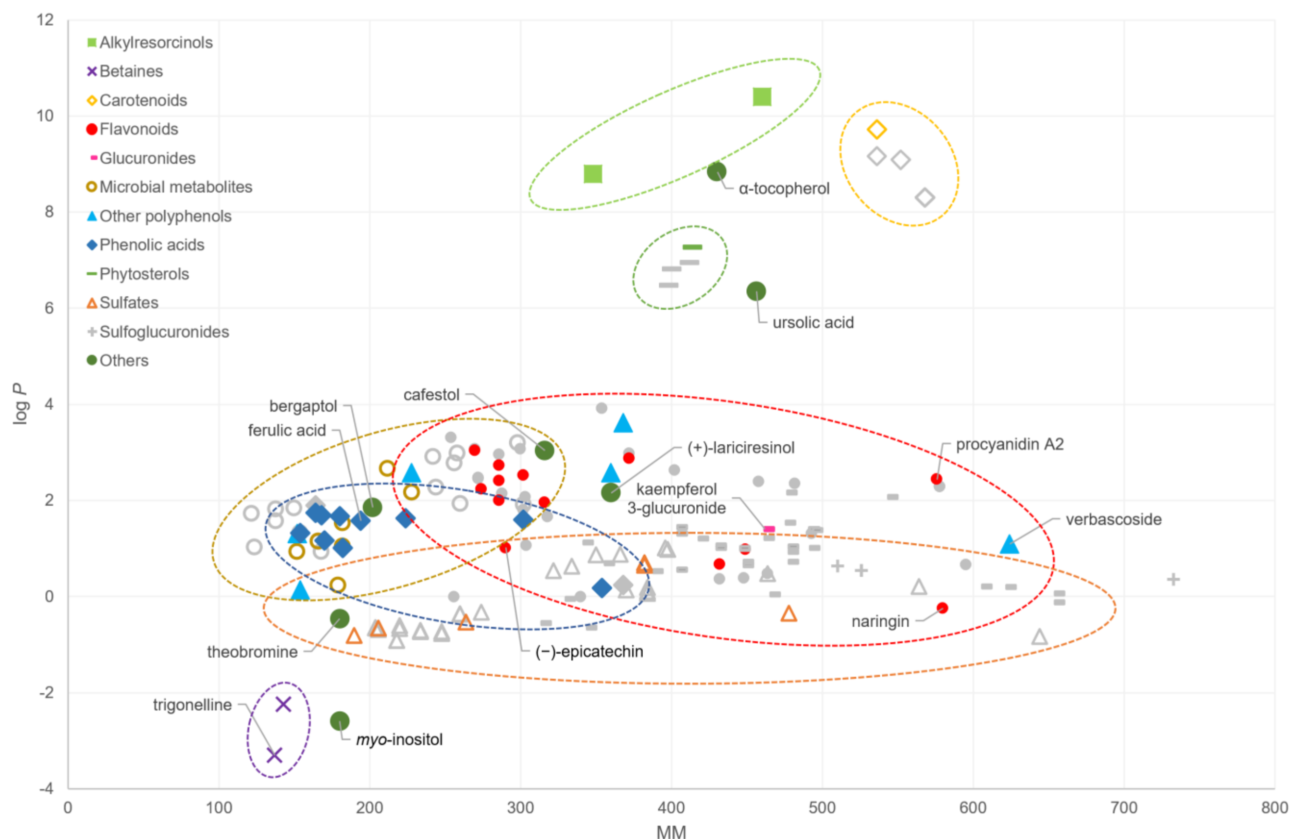


Figure 1. The chemical space (calculated log *P* as a function of monoisotopic molecular mass) of the reference standards included in the analysis (coloured markers, *n* = 56) and included in the initial list (including grey markers, *n* = 179).

Identification of compounds. Identification results were obtained from 10 platforms, including six LC–qTOF platforms, one LC–TOF, one LC–QTRAP, and two GC–MS platforms (Table 1), reflecting the wide use of quadrupole time-of-flight LC–MS instruments in metabolomics analyses. Each platform was capable of detecting and identifying the majority (ranging from 64% to 95%) of the standards from at least one type of mixture (solvent or urine). The highest variability in the rate of detection was compound-specific; Figure 2 shows the identified compounds ranked in the order of positive identifications among both mixture types and ionization modes. Three compounds, bergaptol, genistein and urolithin A, were found in every platform in both mixture types and both ionization modes, followed by ferulic acid, hesperetin and procyanidin A2, which were found in nearly all cases. Most semi-polar compounds, such as flavonoids and phenolic acids, were detected in the LC–MS platforms with relative ease. In contrast, highly lipophilic compounds, such as alkylresorcinols, carotenoids, phytosterols and α-tocopherol, and very hydrophilic

compounds (betaines, sulfated flavonoids, *myo*-inositol and theobromine) were found far less often, most likely due to them being outside of the polarity range of the column used. There is also a possibility that the most lipophilic compounds did not dissolve properly in the selected solvents despite of the visual inspection or they precipitated in further steps of the sample preparation, thus reducing their concentration in the final samples. Among the most elusive standards were also ellagic acid, which has a very low solubility in common solvents (up to 50 ppm) and thus may have been below the limit of detection in several platforms, and cyanidin, which has a low stability and may have deteriorated during the storage of some of the samples. The GC–MS platforms did not detect large (MM > 400) hydrophilic compounds, including flavonoid glycosides and sulfates, as these compounds suffer from low volatility even after TMS derivatization.

- The reliability of identifications: elution order, MS/MS, publicly available databases, figure 3
- Separation of isomers and flavonoids

LC–MS solvent positive | LC–MS solvent negative | LC–MS urine pos. | LC–MS urine neg. | GC–MS

platform 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 | 1 2 4 5 6 7 | 1 2 4 5 6 7 | 9 10 9 10

bergaptol
urolithin A
genistein
ferulic acid
procyanidin A2
hesperetin
urolithin B
apigenin 7-*O*-glucoside
luteolin
vanillic acid
naringin
p-coumaric acid
curcumin
hippuric acid
caffeic acid
protocatechuic acid
vanillin
resveratrol
kaempferol
sinapic acid
4-*O*-methylgallic acid 3-*O*-sulfate
dihydrocaffeic acid
phloretin
(−)-epicatechin
kaempferol 3-glucuronide
isorhamnetin
(+)-lariciresinol
gallic acid
verbascoside
veratric acid
chlorogenic acid
pyrogallol 2-*O*-sulfate
rosmarinic acid
ursolic acid
4-hydroxyphenylacetic acid
homovanillic acid
3-(4-hydroxyphenyl)propionic acid
tangeretin
cyanidin 3-*O*-glucoside
theobromine
myo-inositol
quercetin 3'-*O*-sulfate
cafestol
5-heptadecylresorcinol (AR17:0)
trigonelline
stachydrine (proline betaine)
hydroxytyrosol
cyanidin
quercetin 4'-*O*-sulfate
α-tocopherol
catechol *O*-sulfate
5-pentacosylresorcinol (AR25:0)
ellagic acid
β-sitosterol
β-carotene
quercetin disulfate

Figure 2. The positive identifications (green), unidentified (red) and uncertain identifications (orange) of the chemical standards in different platforms (1 to 10), arranged based on the number of positive identifications. Compounds in light grey were not analyzed. S = identified from solvent, U = identified from urine, + / − = ionization mode; S/N = signal-to-noise ratio < 5, * = overlapping chromatographic peaks, † = detected only in HILIC, ‡ = unexpected retention time and MS/MS.

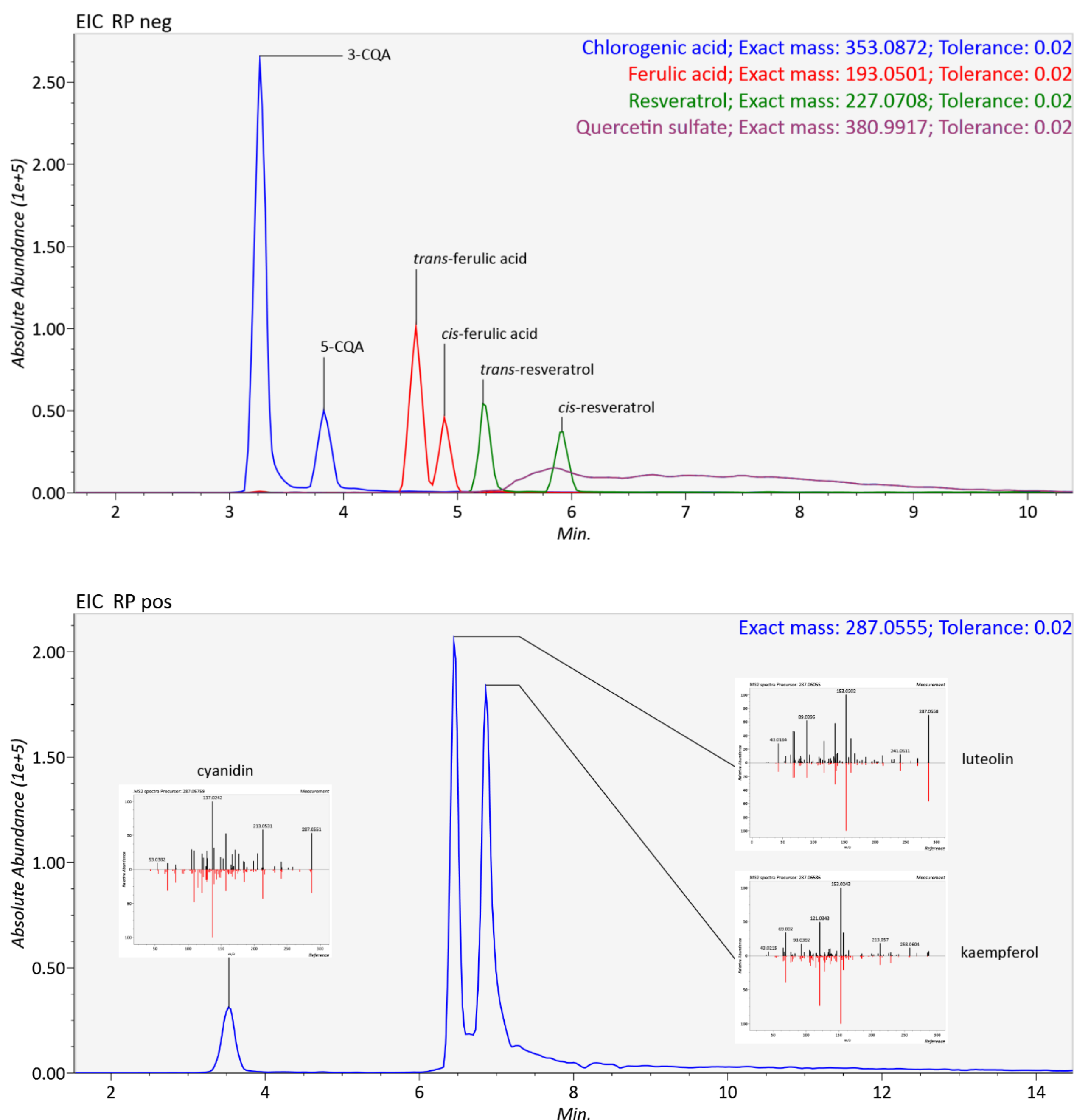


Figure 3. A. Separation of isomers from chlorogenic acid (3-*O*-caffeoylquinic and 5-*O*-caffeoylquinic acid), ferulic acid (*cis* and *trans* isomers), resveratrol (*cis* and *trans* isomers), and quercetin sulfate (quercetin 3-*O*- and 4-*O*-sulfate, from separate chemical standards) in platform 8. In this platform, both forms of chlorogenic acid included in the same reference standard and the *cis* and *trans* isomers of ferulic acid and resveratrol were well separated. However, the isomers of quercetin sulfate were not separated due to extensive tailing. **B.** Separation and identification of three flavonoids, cyanidin, luteolin, and kaempferol, with the same molecular formula of the positive ion ($C_{15}H_{11}O_6^+$) in platform 8. The compounds were identified based on their MS/MS spectra, which were compared with reference spectra from publicly available databases using MS-DIAL software [30].

Influence of chromatography. A variety of column types and lengths (ranging from 50 to 150 mm) was used in the LC–MS platforms (Table 1). All platforms used a reversed-phase

column based on C18 chemistry, and one platform used a complementary HILIC column. The mobile phase was acetonitrile in all platforms except platform 8, where methanol was used instead. Apart from platform 7, where different mobile phases were used for the positive and negative ionization runs, 0.1% formic acid was used as the acidic modifier to increase the ionization of analytes. The length and shape of the gradient, in which the proportion of the organic component of the mobile phase is changed gradually, varied greatly between the platforms (Figure 3). Because of the nearly infinite options to arrange liquid chromatography, the elution of analytes will differ in a way that is challenging to predict. In the case of hesperetin, which was detected in all the platforms, the proportion of the organic solvent did not correlate with the retention times. In the shortest chromatographic run, the proportion reached 100% before the elution of hesperetin, while in the longest run the proportion was 35%. A too short runtime may hinder the differentiation between closely eluting compounds, such as isomers and certain flavonoid species with identical formulas, and should be considered when planning the chromatographic conditions. However, the conditions chosen by each platform, such as the length of the chromatography, did not seem to affect the analytical coverage of the standards in any linear way.

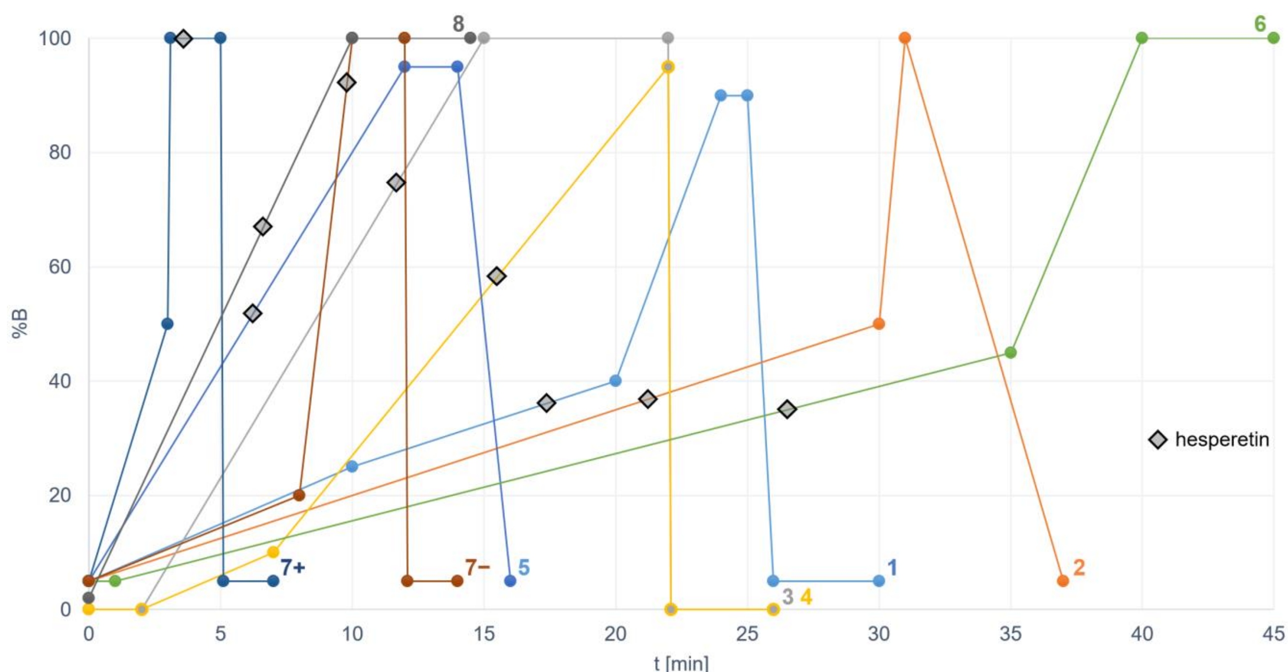


Figure 3. HPLC gradients of the participating LC–MS platforms. The numbers of the platforms correspond with Table 2. Platform no. 7 had a separate HPLC method for positive and negative ionization mode. The retention time of hesperetin in each platform is plotted against the percentage of solvent B.

Ionization mode. The ionization mode in LC–MS often determines whether the analyte will be ionized and detected or not. Phenolic acids possess a carboxylic acid group in their structure, which is prone to deprotonize and obtain a negative charge. However, most

phenolic acids, apart from 3-(4-hydroxyphenyl)propionic acid, were detected in most platforms also in the positive mode. Among all the analyzed standards, only catechol *O*-sulfate and ellagic acid were not detected in the positive mode in any of the platforms. In the negative mode, betaines (due to their inherent positive charge), cafestol, tangeretin and theobromine were not detected. In general, more successful identifications were acquired in the positive mode, suggesting that in the case of only one ionization mode being applied, the positive mode might provide better results for the coverage of plant bioactives in untargeted analyses. Additionally, positive ionization often yields a more extensive fragmentation of the analytes, potentially providing more structural information and reliability in comparing fragmentation spectra with databases [6]. However, there are bias towards certain compound classes or structures when using only one ionization mode, and the negative mode provides important complementary information, such as additional MS/MS spectra and annotation of unknowns. At least for flavonoids, the negative ionization provides highest sensitivity due to lower background noise compared to the positive mode [29].

Conclusions / Considerations on optimal coverage of plant bioactives. The choice of instrumentation is the first obvious factor in determining the analytical coverage of analytes, and it has been widely discussed in the literature. Assuming that the instrument is suitable for metabolomics, there are several other ways to increase the coverage. Sufficient but economical liquid chromatography runtime is essential for the separation of closely eluting peaks. Based on the results presented in this study, a runtime as short as 7 minutes for the positive mode and 14 minutes for the negative mode is adequate for the separation and identification of plant bioactives in simple matrices. The limiting factor in the analysis speed may then be the mass spectrometer's ability to produce spectra at an adequate speed. Since compounds at the extreme ends of the log *P* scale may be out of reach for a single reversed-phase column, the introduction of another complementary column, such as HILIC, will increase the analytical coverage at the hydrophilic end, allowing to choose a reversed-phase column more suitable for lipophilic compounds.

To ensure and validate a wide coverage of plant bioactives in untargeted analyses, we suggest including a simple "analytical coverage quality control mixture" as part of the method optimization. Based on the results presented herein, the mixture would consist of the following four inexpensive and widely available chemical standards:

- (±)-α-tocopherol (e.g. Sigma T3251)
- *trans*-ferulic acid (e.g. Aldrich 128708)
- naringin (e.g. Sigma-Aldrich 91842)
- trigonelline hydrochloride (e.g. Sigma-Aldrich T5509)

The mixture would cover highly lipophilic (α-tocopherol) and very hydrophilic (trigonelline) compounds, compounds with a high molecular mass (naringin) and compounds easily detectable in both positive and negative modes in biological matrices (ferulic acid). It should be noted that only some of the LC–MS platforms in this study were able to detect all four compounds, thus highlighting the importance of such coverage evaluation.

- Concluding remarks

ACKNOWLEDGEMENTS

ITQB/iBET: Conceição Almeida, Ricardo Gomes, Ana Guerreiro, Elsa Mecha

The authors are participating to the COST Action FA 1403 POSITIVe (Interindividual variation in response to consumption of plant food bioactives and determinants involved) funded by the European Union.

REFERENCES

References

1. Allwood, J. W. & Goodacre, R. An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem. Anal.* **21**, 33-47 (2010).

2. Kaspar, S., Peukert, M., Svatos, A., Matros, A. & Mock, H. MALDI- imaging mass spectrometry–An emerging technique in plant biology. *Proteomics* **11**, 1840-1850 (2011).

3. Kueger, S., Steinhauser, D., Willmitzer, L. & Giavalisco, P. High- resolution plant metabolomics: from mass spectral features to metabolites and from whole- cell analysis to subcellular metabolite distributions. *Plant J* **70**, 39-50 (2012).

4. Wu, H. *et al*. Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography–mass spectrometry. *J. Pharm. Biomed. Anal.* **72**, 267-291 (2013).

5. Koistinen, V. M. & Hanhineva, K. Mass Spectrometry-based Analysis of Whole Grain Phytochemicals. *Crit. Rev. Food Sci. Nutr.*, 00-00 (2015).

6. de Villiers, A., Venter, P. & Pasch, H. Recent advances and trends in the liquid-chromatography–mass spectrometry analysis of flavonoids. *Journal of Chromatography A* **1430**, 16-78 (2016).

7. De Vos, R. C. *et al*. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2**, 778-791 (2007).

8. Scalbert, A. *et al*. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **5**, 435-458 (2009).

9. Abad-García, B., Berrueta, L. A., Garmón-Lobato, S., Gallo, B. & Vicente, F. 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 (2009).

10. Mattila, P., Astola, J. & Kumpulainen, J. Determination of flavonoids in plant material by HPLC with diode-array and electro-array detections. *J. Agric. Food Chem.* **48**, 5834-5841 (2000).

11. Wolfender, J., Ndjoko, K. & Hostettmann, K. The potential of LC- NMR in phytochemical analysis. *Phytochem. Anal.* **12**, 2-22 (2001).

12. Cuyckens, F. & Claeys, M. Mass spectrometry in the structural analysis of flavonoids. *J. Mass Spectrom.* **39**, 1-15 (2004).

13. Fossen, T., Andersen, Ø & Markham, K. in *Flavonoids: chemistry, biochemistry and applications* (eds Andersen, Ø & Markham, K.) 37-142 (CRC Press LLC, Boca Raton, 2006).

14. Flores, M. I. A., Romero-González, R., Frenich, A. G. & Vidal, J. L. M. Analysis of phenolic compounds in olive oil by solid-phase extraction and ultra high performance liquid chromatography–tandem mass spectrometry. *Food Chem.* **134**, 2465-2472 (2012).

15. Liu, A. *et al*. Detection, characterization and identification of phenolic acids in Danshen using high-performance liquid chromatography with diode array detection and electrospray ionization mass spectrometry. *Journal of Chromatography A* **1161**, 170-182 (2007).

16. Inbaraj, B. S., Lu, H., Kao, T. & Chen, B. Simultaneous determination of phenolic acids and flavonoids in Lycium barbarum Linnaeus by HPLC–DAD–ESI-MS. *J. Pharm. Biomed. Anal.* **51**, 549-556 (2010).

17. Proestos, C., Boziaris, I., Nychas, G. & Komaitis, M. Analysis of flavonoids and phenolic acids in Greek aromatic plants: Investigation of their antioxidant capacity and antimicrobial activity. *Food Chem.* **95**, 664-671 (2006).

18. Pyrzynska, K. & Biesaga, M. Analysis of phenolic acids and flavonoids in honey. *TrAC Trends in Analytical Chemistry* **28**, 893-902 (2009).

19. Naczk, M. & Shahidi, F. Phenolics in cereals, fruits and vegetables: Occurrence, extraction and analysis. *J. Pharm. Biomed. Anal.* **41**, 1523-1542 (2006).

20. Willför, S., Smeds, A. & Holmbom, B. Chromatographic analysis of lignans. *J. Chromatogr. A* **1112**, 64-77 (2006).

21. Abidi, S. Chromatographic analysis of plant sterols in foods and vegetable oils. *Journal of chromatography A* **935**, 173-201 (2001).

22. Lerma-García, M. J., Simó-Alfonso, E. F., Méndez, A., Lliberia, J. L. & Herrero-Martínez, J. M. Fast separation and determination of sterols in vegetable oils by ultraperformance liquid chromatography with atmospheric pressure chemical ionization mass spectrometry detection. *J. Agric. Food Chem.* **58**, 2771-2776 (2010).

23. Urpi-Sarda, M. *et al*. Targeted metabolic profiling of phenolics in urine and plasma after regular consumption of cocoa by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A* **1216**, 7258-7267 (2009).

24. Sang, S., Lee, M., Yang, I., Buckley, B. & Yang, C. S. Human urinary metabolite profile of tea polyphenols analyzed by liquid chromatography/electrospray ionization tandem mass spectrometry with data- dependent acquisition. *Rapid Communications in Mass Spectrometry* **22**, 1567-1578 (2008).

25. Ross, A. B., Kamal-Eldin, A. & Aman, P. Dietary alkylresorcinols: absorption, bioactivities, and possible use as biomarkers of whole-grain wheat- and rye-rich foods. *Nutr. Rev.* **62**, 81-95 (2004).

26. Koistinen, V. M., Katina, K., Nordlund, E., Poutanen, K. & Hanhineva, K. Changes in the phytochemical profile of rye bran induced by enzymatic bioprocessing and sourdough fermentation. *Food Res. Int.* (2016).

27. Sánchez- Hernández, L., Castro- Puyana, M., Luisa Marina, M. & Crego, A. L. Determination of betaines in vegetable oils by capillary electrophoresis tandem mass spectrometry–application to the detection of olive oil adulteration with seed oils. *Electrophoresis* **32**, 1394-1401 (2011).

28. Guo, Y. Analysis of quaternary amine compounds by hydrophilic interaction chromatography/mass spectrometry (HILIC/MS). *J. Liq. Chromatogr. Rel. Technol.* **28**, 497-512 (2005).

29. de Rijke, E. *et al*. Analytical separation and detection methods for flavonoids. *Journal of Chromatography A* **1112**, 31-63 (2006).

30. Tsugawa, H. *et al*. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* **12**, 523-526 (2015).

FIGURES

(move here before submission)

TABLES

Table 1. Platforms that participated in the test with their equipment information and chromatographic conditions. ESI = electrospray ionization, EI = electron ionization. To be removed from the submitted manuscript: Platform 1 = CEBAS, 2 = ILVO, 3 = INRA plasma method, 4 = INRA urine method, 5 = ITQB, 6 = SZIU, 7 = UB, 8 = UEF, 9 = SLU, 10 = VTT

Platform	General method(s)	HPLC/GC model	Column	MS	Ion source	HPLC mobile phase	HPLC flow	HPLC gradient (t [min], %B)
1	UPLC–QTOF	Agilent 1290	Agilent Poroshell 120 EC-C18 (3 × 100 mm, 2.7 µm)	Agilent 6550	ESI+/–	A: H_2O + 0.1% FA, B: ACN + 0.1% FA	0.4 ml/min	(0, 5), (10, 25), (20, 40), (24, 90), (25, 90), (26, 5), (30, 5)
2	UPLC–TOF	Acquity H-class	Acquity UPLC BEH Shield RP18 (2.1 × 150 mm, 1.7 µm)	Synapt G2 S	ESI+/–	A: H_2O + 0.1% FA, B: ACN + 0.1% FA	0.35 ml/min	(0, 5), (30, 50), (31, 100), (37, 5)
3	HPLC–QTOF	Thermo U3000	Acquity HSS T3, 100 Å (2.1 × 150 mm, 1.8 µm)	Bruker Impact HD2	ESI+/–	A: H_2O + 0.1% FA, B: ACN + 0.1% FA	0.4 ml/min	(0, 0), (2, 0), (15, 100), (22, 100), (22.1, 0), (26, 0)
4	HPLC–QTOF	Thermo U3000	Acquity UPLC BEH Shield RP18 (2.1 × 100 mm, 1.7 µm)	Bruker Impact HD2	ESI+/–	A: H_2O + 0.1% FA, B: ACN + 0.1% FA	0.4 ml/min	(0, 0), (2, 0), (7, 10), (22, 95), (22.1, 0), (26, 0)
5	UPLC–QTOF	Eksigent nanoLC	Eksigent HALO C18, 90 Å (0.5 × 50 mm, 2.7 µm)	Sciex TripleTOF 6600	ESI+/–	A: H_2O + 0.1 % FA, B: ACN + 0.1% FA	10 µl/min	(0, 5), (12, 95), (14, 95), (16, 5)
6	UPLC–QTOF	Agilent 1260	Phenomenex Synergi Hydro-RP (150 × 2 mm, 4 µm)	Agilent 6543	ESI+/–	A: H_2O + 0.1% FA, B: ACN + 0.1% FA	0.5 ml/min	(0, 5), (1, 5), (35, 45), (40, 100), (45, 100)
7	UPLC−QTRAP	Agilent 1290	Luna Omega Polar C18 (100 × 2.1 mm, 1.6 µm)	Sciex 6500	ESI+	A: H_2O + 0.5% FA, B: ACN + 0.5% FA	0.5 ml/min	(0, 5), (3, 50), (3.1, 100), (5, 100), (5.1, 5), (7, 5)
7	UPLC−QTRAP	Agilent 1290	Luna Omega Polar C18 (100 × 2.1 mm, 1.6 µm)	Sciex 6500	ESI–	A: H_2O + 0.1% FA + 10 mM NH_4COOH, B: ACN	0.5 ml/min	(0, 5), (8, 20), (10, 100), (12, 100), (12.1, 5), (14, 5)
8	UPLC–QTOF	Agilent 1290	Agilent Zorbax Eclipse XDB-C18 (2.1 × 100 mm, 1.8 µm) + Aqcuity UPLC BEH amide HILIC (2.1 × 100 mm, 1.7 µm)	Agilent 6540	ESI+/–	A: H_2O + 0.1% FA, B: MeOH + 0.1% FA	0.4 ml/min	(0, 2), (10, 100), (14.5, 100), (14.51, 2), (16.5, 2)
9	GC–MS					–		–
10	GC–MS	Agilent 7890A	DB-5ms (30 m, 0.25 mm, 0.25 µm)	Agilent 5975C MSD	EI	–		–

Supplementary material

Supplementary Table n. Chemical standards analysed in the multiplatform test. Concentration in the stock solution other than 10 mM: * 5 mM, † 1 mM, ‡ 0.5 mM, § 50 ppm

Compound	Class	Formula	M_0 [Da]	log *P* (calc.)	Supplier	Mix
trigonelline	betaine	$C_7H_7NO_2$	137.0477	−3.30	Extrasynthese	A
stachydrine (proline betaine)	betaine	$C_7H_{13}NO_2$	143.0946	−2.24	Extrasynthese	A
4-hydroxyphenylacetic acid	microbial metab.	$C_8H_8O_3$	152.0473	0.93	Aldrich	A
vanillin	other phenolic	$C_8H_8O_3$	152.0473	1.31	Sigma-Aldrich	A
protocatechuic acid	phenolic acid	$C_7H_6O_4$	154.0266	1.32	Sigma	A
hydroxytyrosol	other phenolic	$C_8H_{10}O_3$	154.0630	0.13	Extrasynthese	B
p-coumaric acid	phenolic acid	$C_9H_8O_3$	164.0473	1.74	Sigma	A
3-(4-hydroxyphenyl)propionic acid	microbial metab.	$C_9H_{10}O_3$	166.0630	1.15	Fluka	A
vanillic acid	phenolic acid	$C_8H_8O_4$	168.0423	1.70	Sigma	A
gallic acid	phenolic acid	$C_7H_6O_5$	170.0215	1.17	Sigma (G73849	A
hippuric acid	phenolic acid	$C_9H_9NO_3$	179.0582	0.23	Sigma (112003)	A
caffeic acid	phenolic acid	$C_9H_8O_4$	180.0423	1.67	Sigma (C0625)	A
myo-inositol	sugar alcohol	$C_6H_{12}O_6$	180.0634	−2.59	Merck	A
theobromine	alkaloid	$C_7H_8N_4O_2$	180.0650	−0.46	Extrasynthese	A
dihydrocaffeic acid	microbial metab.	$C_9H_{10}O_4$	182.0579	1.04	Sigma	A
homovanillic acid	phenolic acid	$C_9H_{10}O_4$	182.0579	1.02	Extrasynthese	B
veratric acid	phenolic acid	$C_9H_{10}O_4$	182.0579	1.52	Aldrich	A
catechol-*O*-sulfate	sulfate	$C_6H_6O_5S$	189.9936	−0.81	*synthesised in-house*	B†
ferulic acid	phenolic acid	$C_{10}H_{10}O_4$	194.0579	1.58	Aldrich (128708)	A
bergaptol	other phenolic	$C_{11}H_6O_4$	202.0270	1.86	Extrasynthese	B*
pyrogallol-2-*O*-sulfate	sulfate	$C_6H_6O_6S$	205.9885	−0.66	*synthesised in-house*	A
urolithin B	microbial metab.	$C_{13}H_8O_3$	212.0473	2.65	n/a	B
sinapic acid	phenolic acid	$C_{11}H_{12}O_5$	224.0685	1.63	Sigma	A†
urolithin A	microbial metab.	$C_{13}H_8O_4$	228.0423	2.16	n/a	B
resveratrol	other phenolic	$C_{14}H_{12}O_3$	228.0786	2.57	Sigma	A
4-*O*-methylgallic acid 3-*O*-sulfate	sulfate	$C_8H_8O_8S$	263.9940	−0.53	*synthesised in-house*	A
genistein	flavonoid	$C_{15}H_{10}O_5$	270.0528	3.04	Extrasynthese	B
phloretin	other phenolic	$C_{15}H_{14}O_5$	274.0841	2.23	Extrasynthese	B
kaempferol	flavonoid	$C_{15}H_{10}O_6$	286.0477	1.99	Extrasynthese	B
luteolin	flavonoid	$C_{15}H_{10}O_6$	286.0477	2.73	Fluka	B*
cyanidin	flavonoid	$C_{15}H_{11}O_6^+$	287.0556	2.41	n/a	B
(−)-epicatechin	flavonoid	$C_{15}H_{14}O_6$	290.0790	1.02	Sigma	B
ellagic acid	phenolic acid	$C_{14}H_6O_8$	302.0063	1.59	Sigma	A§
hesperetin	flavonoid	$C_{16}H_{14}O_6$	302.0790	2.52	Sigma	A
isorhamnetin	flavonoid	$C_{16}H_{12}O_7$	316.0583	1.96	Extrasynthese	B‡
cafestol	terpenoid	$C_{20}H_{28}O_3$	316.2038	3.04	MP Biomedicals	A

5-heptadecylresorcinol (AR17:0)	alkylresorcinol	$C_{23}H_{40}O_2$	348.3028	8.79	ReseaChem	B
chlorogenic acid	phenolic acid	$C_{16}H_{18}O_9$	354.0950	0.17	Aldrich (C3878)	A
rosmarinic acid	phenolic acid	$C_{18}H_{16}O_8$	360.0845	2.57	Extrasynthese	A
(+)-lariciresinol	lignan	$C_{20}H_{24}O_6$	360.1573	2.16	ArboNova	A
curcumin	other phenolic	$C_{21}H_{20}O_6$	368.1260	3.62	Sigma	B*
tangeretin	flavonoid	$C_{20}H_{20}O_7$	372.1210	2.88	Extrasynthese	B
quercetin 3'-*O*-sulfate	sulfate	$C_{15}H_{10}O_{10}S$	381.9995	0.65	*synthesised in-house*	A†
quercetin 4'-*O*-sulfate	sulfate	$C_{15}H_{10}O_{10}S$	381.9995	0.70	*synthesised in-house*	A†
β-sitosterol	steroid	$C_{29}H_{50}O$	414.3862	7.27	Sigma	C
α-tocopherol	other phenolic	$C_{29}H_{50}O_2$	430.3811	8.84	Sigma	A
apigenin 7-*O*-glucoside	flavonoid	$C_{21}H_{20}O_{10}$	432.1056	0.68	HWI Analytik	B*
cyanidin 3-*O*-glucoside	flavonoid	$C_{21}H_{21}O_{11}^+$	448.1006	0.98	*synthesised in-house*	B
ursolic acid	terpenoid	$C_{30}H_{48}O_3$	456.3603	6.35	Aldrich	A
5-pentacosylresorcinol (AR25:0)	alkylresorcinol	$C_{31}H_{56}O_2$	460.4280	10.4	ReseaChem	B*
kaempferol 3-glucuronide	glucuronide	$C_{21}H_{18}O_{12}$	462.0798	1.40	*synthesised in-house*	A
quercetin disulfate	sulfate	$C_{15}H_{10}O_{13}S_2$	461.9563	−0.34	*synthesised in-house*	B‡
β-carotene	terpenoid	$C_{40}H_{56}$	536.4382	9.72	Extrasynthese	C
procyanidin A2	flavonoid	$C_{30}H_{24}O_{12}$	576.1268	2.43	Extrasynthese	A
naringin	flavonoid	$C_{27}H_{32}O_{14}$	580.1792	−0.24	Sigma	A
verbascoside	phenolic acid	$C_{29}H_{36}O_{15}$	624.2054	1.09	Extrasynthese	A

Supplementary Table n. A list of 162 plant bioactives and their metabolites selected as candidates for the multiplatform analysis classified based on their chemical or metabolite class.

Carotenoids

Lutein
Lycopene
Zeaxanthin
β-Carotene
β-Cryptoxanthin

Phytosterols

Brassicasterol
Campesterol
Stigmasterol
β-Sitosterol

Phenolic acids

2-Hydroxycinnamic acid
5-Caffeoylquinic acid
5-Feruloylquinic acid
Caffeic acid
Ellagic acid
Ferulic acid

Isorhamnetin 3-O-glucuronide
Isorhamnetin 4'-O-glucuronide
Isovanillic acid 3-O-glucuronide
Kaempferol 3-O-β-D-glucuronide
Luteolin 7-O-β-D-glucuronide
Myricetin 3'-O-glucuronide
Naringenin 4'-O-β-D-glucuronide
Naringenin 7-O-β-D-glucuronide
Protocatechuic acid 4-O-glucuronide
Quercetin 3'-O-β-D-glucuronide
Quercetin 5,7-diglucuronide
Quercetin 7-O-β-D-glucuronide
trans-Resveratrol 3-O-β-D-glucuronide
trans-Resveratrol 4'-O-β-D-glucuronide
Vanillic acid 4-O-glucuronide

Sulfates

1-Methylpyrogallol 3-O-sulphate
2-Methylpyrogallol 1-O-sulphate

Gallic acid
Homovanillic acid
Sinapic acid
Vanillic acid
Verbascoside

Flavonoids

(-)-Epicatechin
(-)-Epigallocatechin 3-O-gallate
8-Prenylnaringenin
Apigenin
Cyanidin
Cyanidin 3-O-rutinoside
Daidzein
Delphinidin
Dihydrogenistein
Diosmetin
Eriodictyol
Genistein
Hesperetin
Homoeriodictyol
Homoorientin
Isoliquiritigenin
Isoquercitrin
Isorhamnetin
Isosakuranetin
Kaempferol
Luteolin
Malvidin 3-O-glucoside
Myricetin
Naringenin
Nobiletin
Phloretin
Quercetin
Silybin
Sinensetin
Tangeretin
Taxifolin
Vitexin
Xanthohumol

Procyanidins

Procyanidin dimer A2
Procyanidin dimer B1
Procyanidin dimer B2
Procyanidin dimer B4

Other polyphenols

Curcumin
Hydroxytyrosol
Resveratrol

Glucuronides

3,4-Dihydroxybenzoic acid 3-O-sulfate
3,4-Dihydroxybenzoic acid 4-O-sulfate
3'-O-Methyl-(-)-epicatechin 5-O-sulfate
3'-O-Methyl-(-)-epicatechin 7-O-sulfate
4-Hydroxybenzoic acid 4-O-sulphate
4-Methylcatechol 1-O-sulphate
4-Methylcatechol 2-O-sulphate
4-O-Methylgallic acid 3-O-sulphate
Benzoic acid sulfate
Caffeic acid 3-O-sulphate
Caffeic acid 4-O-sulphate
Catechol O-sulfate
Daidzein 4' sulfate
Epicatechin 3-O-sulfate
Genistein 7-sulfate
Isoferulic acid 3-O-sulfate
Isoquercitrin 4'-O-sulfate
Isorhamnetin 3-O-sulfate
Isovanillic acid 3-O-sulfate
Kaempferol 3-O-sulfate
Myricetin 3'-O-sulfate
Protocatechuic acid 3-O-sulphate
Protocatechuic acid 4-O-sulphate
Pyrogallol 1-O-sulfate
Quercetin 3'-O-sulfate
Quercetin 3-O-sulfate
Quercetin 4'-O-sulfate
Silybin 20-O-sulfate
Silybin 7,20-di-O-sulfate
Taxifolin 4'-O-sulfate
Vanillic acid 4-O-sulphate

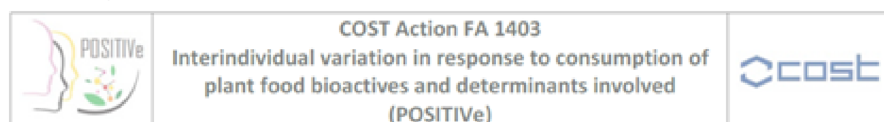
Sulfates and glucuronides

3'-Methylcyanidin 3-glucuronide-5-glucoside-4'-sulfate
Daidzein 7-β-D-glucuronide 4’-Sulfate
Genistein 7-sulfate 4’-β-D-glucuronide
Genistein 7-β-D-glucuronide 4’-sulfate

Microbial metabolites

3,4-Dihydroxybenzoic acid
3,4-Dihydroxyphenylacetic acid
3,4-Dihydroxyphenylpropionic acid
3,4-Dimethoxybenzoic acid
3-Hydroxybenzoic acid
3-Hydroxyphenylacetic acid
3-Hydroxyphenylpropionic acid
3-Phenylpropionic acid
4-Coumaric acid
4-Hydroxybenzoic acid
4-Hydroxyphenylacetic acid
4-Hydroxyphenylpropionic acid

3'-Methylcyanidin 3-glucuronide
4-Hydroxybenzoic acid 4-O-glucuronide
Apigenin 7-glucuronide
Caffeic acid 3-O-β-D-glucuronide
cis-Resveratrol 3-O-β-D-glucuronide
cis-Resveratrol 4'-O-β-D-glucuronide
Curcumin-4-O-β-D-glucuronide
Cyanidin 3-glucuronide
Daidzein 4’-β-D-glucuronide
Daidzein 7-β-D-glucuronide
Daidzein diglucuronide
Diosmetin 3'-O-β-D-glucuronide
Epicatechin 3-O-glucuronide
Genistein 4’-β-D-glucuronide
Genistein 7-β-D-glucuronide
Genistein diglucuronide
Hesperetin 3'-O-β-D-glucuronide
Hesperetin 7,3'-di-O-β-D-glucuronide
Hesperetin 7-O-β-D-glucuronide
4-Methylcatechol
Benzoic acid
Dihydrocaffeic acid
Dihydrodaidzein
Enterodiol
Enterolactone
Equol
Equol 4'-sulfate
Equol 7-β-D-glucuronide
Hippuric acid
O-Desmethylangolensin
Protocatechuic acid
Urolithin A
Urolithin A 3-glucuronide
Urolithin A 8-glucuronide
Urolithin B
Urolithin B 3-O-glucuronide
Urolithin C
Urolithin D



Contents

1. AIM 3
2. DOCUMENTS GENERATED BY THIS SOP 3
3. PROCEDURE 4
3.1. Preparation of stock solutions 4
3.1.1. Compounds to include in the coverage test 4
3.1.2. Collection of compounds to include in the ring test 4
3.1.3. Preparation of the individual standard stock solutions at 10mM 4
3.1.4. Preparation of the standard mix solution at 200μM (S1) 4
3.1.5. Preparation of the standard stock solution at 10 μM (S2 and U2) 5
3.2. Analysis of standard solutions S2 and U2 5
3.3. Treatment of results 6
Annex 1. List of Analytical platforms envolved and contact persons 7
Annex 2. List of compounds, formula, monoisotopic mass (MM), log P, origin, purity, solubility 8
Annex 3. Characteristics of the MS equipment and conditions of analysis 9
Annex 4. Characteristics of the GC-MS equipment and conditions of analysis 12
Annex 5. Characteristics of the NMR equipment and conditions of analysis 13
Annex 6. Template to provide results from each platform 14

Supplementary Figure 1. The contents of the standard operating procedure (SOP) sent to each participating platform.