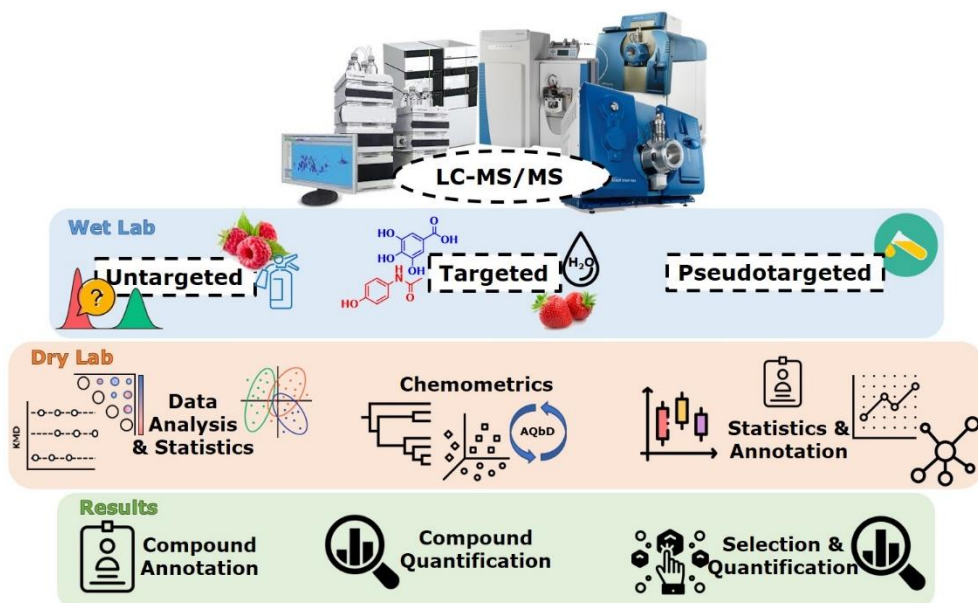# PhD IN CHEMICAL SCIENCES

CYCLE XXXV

COORDINATOR Prof. ANNA MARIA PAPINI

DEVELOPMENT AND APPLICATION OF INSTRUMENTAL LC-MS/MS

ANALYTICAL PLATFORMS AND CHEMOMETRIC DATA TREATMENT

FOR THE STUDY OF COMPLEX MATRICES IN THE

ENVIRONMENTAL AND NUTRITIONAL FIELDS

PhD Candidate	**Tutor**
Dr. Lapo Renai	Prof. Massimo Del Bubba

PhD IN CHEMICAL SCIENCES
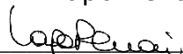
CYCLE XXXV

COORDINATOR Prof. ANNA MARIA PAPINI

DEVELOPMENT AND APPLICATION OF INSTRUMENTAL LC-MS/MS

ANALYTICAL PLATFORMS AND CHEMOMETRIC DATA TREATMENT FOR

THE STUDY OF COMPLEX MATRICES IN THE ENVIRONMENTAL AND

NUTRITIONAL FIELDS

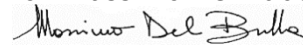Academic Discipline CHIM/01

PhD Candidate   **Tutor**

Dr. Lapo Renai   Prof. Massimo Del Bubba

Coordinator

Prof. Anna Maria Papini

Academic Years 2019/2022

Index

Abstract

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) represents a fundamental analytical tool for the in-depth characterization of complex matrices, such as environmental samples, foods, and biological fluids. In fact, reliable targeted and untargeted LC-MS/MS platforms are of paramount importance in the aforementioned fields for the identification and quantification of micropollutants and their transformation products in environmental samples, as well as food constituents and human metabolites in metabolomics studies. Also, the use of proper chemometrics and data analysis tools is crucial in both method development and data exploration/interpretation, the latter especially in untargeted studies. This PhD project has been carried out to develop high throughput LC-MS/MS platforms exploiting both targeted and untargeted approaches capable to untangle the intricate data information from complex samples of environmental and nutritional interest. To this extent, the latest advanced wet (e.g., on-line SPE and LC×LC) and dry (e.g., molecular networking) lab platforms have been employed in the development of LC-MS/MS methods to investigate different types of (analytically challenging) samples, including fruit extracts, wastewaters, technical mixtures, and human biofluids (i.e., urine and serum). The LC systems used for the development of untargeted and targeted platforms were configured to perform: (i) low- (tens µL) and large-volume direct injection (hundreds µL), (ii) on-line SPE pre-treatment, and (iii) LC×LC analyses. Targeted experiments are performed using a hybrid triple quadrupole analyzer operating in MRM mode, capable to provide high sensitivity, accuracy and robustness required in quantitative protocols. The untargeted studies were conducted using two different high-resolution MS/MS analyzers, based on TOF (separation "in space") and Orbitrap (separation "in time") technologies, allowing for performing accurate mass determination (<1-10 ppm), isotopic profiling for molecular formula assignment, and tandem mass experiments for structural elucidation of unknow compounds. The development and/or application of (i) LC-MS/MS wet lab methods and (ii) dry lab chemometric workflows presented in this

PhD thesis generated highly informative and accurate data that could help to increase the knowledge on the chemical space of the investigated samples. Finally, when applicable, LC-MS/MS data have been publicly shared in available repositories (e.g., MassIVE) to contribute to data dissemination and usability within the Scientific Community.

1   General introduction

1.1 Challenges in liquid chromatography-tandem mass spectrometry (LC-MS/MS)

The progresses in analytical instrumentation allowed for increasing the selectivity and sensitivity performances of modern liquid chromatography (LC, both high and "ultra-high" performances) hyphenated with tandem mass spectrometry (MS/MS) platforms, pushing the boundaries of the detectable chemical space [1].

The first key point in the analytical improvement of LC instruments is represented by the reduction of particle size distribution (PSD) in column packing of fully porous particles from 5 μm to <2 μm, which allows for reaching the same number of theoretical plates in a very short time of analysis, and thus determining a gain in terms of chromatographic efficiency (i.e., ultra-high performances LC) [2]. In parallel, LC hardware evolved to deliver higher operating pressures to guarantee high kinetic performances, since mobile-phase linear velocity also increases inversely proportional to PSD. In this regard, the introduction of core-shell particle technology improved significantly chromatographic performances, limiting flow resistance (i.e., system backpressure) and providing reduced plate height [3]. The technical progress of stationary phase supports goes hand-in-hand with the column partition chemistry. Traditional reversed stationary phases (RP), like alkyl silica-bonded (e.g., C18 and C8) phases, now include novel hydrophobic functionalities capable to exploit more than a single partition mechanism, such as pi-pi (e.g., phenyl-hexyl and biphenyl functions), dipole-dipole (e.g., pentafluorophenyl functions), and weak or strong ionic exchange (anionic or cationic) interactions, the latter known also as mixed-mode LC (MMLC) [4]. The extension of the investigable chemical space due to both the introduction of innovative RPLC stationary phases and the improvement of size exclusion (e.g., size exclusion chromatography - SEC) and normal phase like (e.g., hydrophilic interaction chromatography - HILIC) separation mechanisms, has promoted the interest in on-line bidimensional chromatographic systems (e.g., 2D-LC and LC×LC) combining different selectivity processes capable to increase the system

peak capacity (i.e., the number of separable components), especially when very complex samples need to be characterized [5].

The technological improvement in chromatographic techniques not only allows to increase the selectivity (e.g., chemical space) and peak coverage (e.g., elution, shape, and resolution) of the chromatographic domain, but advantages are also gained in the mass domain. In this regard, the combination of more effective atmospheric pressure ionization (API) devices, commonly electrospray ionization (ESI), and of advanced MS/MS and high-resolution MS (HRMS) analyzers with the enhanced chromatographic selectivity and resolution, allows for providing reliable data for (i) structure elucidation of unknown compounds (i.e., annotation) and (ii) quantitative analysis [6]. Depending on the type of LC-MS/MS platform used and the data structure obtained, it is now common to define untargeted (or non-targeted) analysis the comprehensive profiling of known and unknown compounds occurring in a sample, whereas when the focus is made on a group of known chemicals, the set definition is targeted analysis [7]. In addition, due to the different data structure obtained with the two approaches, post-run data handling and processing deeply differ between untargeted (e.g., statistical-based feature detection) and targeted (e.g., least squares linear regressions) approaches. Recently, the pseudotargeted workflow has been introduced, in which unknown compounds are at first annotated by an untargeted method and then quantified using a targeted approach [4].

A deeper investigation of the instrumental and data-processing features for untargeted and targeted LC-MS/MS platforms is reported in the following section.

1.2 LC-MS/MS-based analytical platforms

1.2.1 Untargeted LC-MS/MS

The most recent untargeted platforms combine LC systems with hybrid HRMS analyzers, such as quadrupole-time of flight (Q-ToF) and quadrupole-Orbitrap configurations. The high resolution (from 20,000 to ≥100,000) provided by these instruments, also due to the hybridization with quadrupoles and/or linear ion traps

that select a limited number of ions to be sent to the analyzer, allows to perform accurate mass measurements (m/z) of the precursor ions. (or pseudo-molecular ions) and fragmentation products in MS/MS experiments. These traits allow for annotating unknown compounds by (i) accurate mass determination (e.g., 10 ppm or lower) and isotopic profile analysis of precursor ions for molecular formula assignment, and (ii) mass fragmentation patterns (MS/MS or MS^n) for structure elucidation [8, 9]. The continuous hardware and software development has allowed the implementation of mixed acquisition techniques compared to the traditional full scan, namely data-dependent acquisition (DDA) and data independent acquisition (DIA) [10].

Despite the wide range of achievable data and information, the choice of a proper LC configuration is mandatory even in untargeted platforms, since it exerts a significant influence on peak coverage (i.e., stationary phase selectivity) and sensitivity (i.e., mobile phase composition) of the method. Once retention times, m/z (i.e., a feature), and fragments are extracted, feature annotation can be confirmed by matching reference standards (level I, i.e., MS data and retention) or putatively identified (levels II-III) in comparison with reference spectral libraries (e.g., mzCloud, Human Metabolome Data Base etc.). In the latter case, spectral data can be compared by manual investigation or by automatic querying within commercial (e.g., Compound Discoverer and Marker View) or publicly available softwares (e.g., MZmine, MS-Dial, GNPS etc.) [11]. Additionally, to reduce the workload of the annotation process, only the more interesting features, according to the experimental design and statistical analysis, can be selected and identified [12]. Alternatively, the detected features are checked within an *a priori* list of compounds of interest (i.e., a suspect list), also known as suspect screening approach [13].

Given the comprehensiveness of the obtainable information, the untargeted LC-MS/MS platforms have found application and wide development in metabolomics, lipidomics, and proteomics [6, 14]. Recently, untargeted profiling has been also applied to the analysis of contaminants of emerging concern and

their transformation products in environmental samples [13], as well as for the identification of their biomarker of human exposure (i.e., exposomics) [10]. Furthermore, the growing awareness on sample complexity in omics applications has increased the focus on untargeted 2D-LC platforms, which provide enhanced resolution and greater chemical space coverage, thanks also to technological progresses in API coupling with HRMS, reducing the potential incompatibilities related to the composition and flow of the mobile phases (e.g., RPLC-HILIC) [15].

1.2.2   Targeted LC-MS/MS

Targeted platforms foresee the coupling of the LC system with both HRMS analyzers and unit resolution analyzers, such as triple quadrupole mass spectrometers (QqQ). The essential requirement is to set up the tandem mass experiments to follow certain product ions common to one or more precursor ions. Examples of these acquisition strategies are (i) parallel reaction monitoring (PRM) [16] and (ii) selected/multiple reaction monitoring (SRM/MRM) [17], in HRMS and unit resolution MS, respectively. Of particular interest for targeted applications are the QqQ analyzers, due to the high transmission and scan rate given by the quadrupolar geometry and by the application of cell potentials and radiofrequencies, providing high sensitivity in a wide range of m/z (e.g., 1-2000 m/z). For these reasons, targeted experiments are applied to quantitative analysis protocols, with the only limit of analytical standard availability, for which it is mandatory to perform accurate calibration curves (external or matrix-matched calibrations) of the instrumental response (e.g., peak area) as a function of the concentration of the selected/available compounds [18]. The use and/or scouting of stationary and elution gradients is of analytical relevance in this context, since it allows for obtaining the best chromatographic resolution, thus solving the issues of (i) isobaric compounds identification/quantification and (ii) method sensitivity in case of the elution of many analytes in a narrow retention time window [19].

The quality of the quantitative data is confirmed by validation protocols including the evaluation of sensitivity, precision, and trueness of the targeted method, the latter given by matrix effect (ME) and recovery (R%) evaluation. ME refers to the suppression or enhancement of analyte signal during the ionization process by interfering compounds in the sample matrix, whereas R% expresses the efficiency of the sample preparation and/or extraction steps, when required [18]. The inclusion in the targeted platforms of both off-line (e.g., SPE, QuEChERS, d-SPE etc.) and on-line (e.g., on-line SPE) sample extraction procedures allows to increase the sensitivity and the R% of the method, helping to reduce the ME by removing matrix co-extracted constituents during sample processing [20, 21]. This wide range of quantitative outputs paves the way to the use of numerous chemometric strategies for data processing, suitable for both the extraction of latent information for sample discrimination/classification (e.g., uni- and multivariate analyses) [7], but also for the optimization of method validation parameters (e.g., ME and R) through multivariate approaches (e.g., design of experiment and quality by design) [22].

Given the exploitable chemical space and the possibility of obtaining accurate quantitative data, the LC-MS/MS targeted platforms are used in numerous fields, such as food safety, foodomics, environmental safety, and clinical analysis [23-26].

1.3 Environmental and nutritional applications: thesis background

The untargeted and targeted LC-MS/MS platforms described in paragraphs above and their specific development and application to foodomics, nutritional metabolomics (i.e., nutrimetabolomics), and environmental studies, laid the ground of this thesis work.

Untargeted LC-HRMS and LC-MS/MS for accurate mass readout and structure elucidation was adopted for (i) the comprehensive (poly)phenolic profiling of *Vaccinium myrtillus*, *Vaccinium corymbosum*, and *Vaccinium uliginosum* L. subsp. *gaultherioides* berries [27], and the (ii) identification of (poly)phenol

metabolites in human serum and urine after acute ingestion of a *Vaccinium myrtillus* berry supplement [28].

In the first study, several secondary metabolites belonging to the classes of anthocyanins, monomeric and oligomeric flavonols, flavanols, dihydrochalcones, phenolic acids, together with other structurally mixed (poly)phenolic compounds (e.g., A/B-type proanthocyanidins), were identified at levels I and II among the investigated berries. Additionally, their (poly)phenolic composition successfully discriminated the three *Vaccinium* berry species, underlying the relevant role of untargeted LC-MS/MS platforms for foodomics applications [27]. The acquired knowledge on the (poly)phenolic profile of *Vaccinium* berries, allowed for investigating the metabolic structural alterations of (poly)phenols upon human intake through a two-arm intervention study of *Vaccinium myrtillus* and *Vaccinium corymbosum*. A first investigation on human serum and urine after the intake of *Vaccinium myrtillus* berry highlighted the presence of thirty-six (poly)phenolic derivatives, such as benzoic acids, hydroxyhippuric acids, cinnamic acids, phenylpropionic acids, phenylvaleric acids, phenylpentenoic acids and abscisic acid, which were putatively identified in the investigated biofluids by XCMS-minfrac data processing and statistical analysis [28].

Targeted LC-MS/MS platforms play a key role in foodomics too, since the chance to select compounds covering the chemical space of interest and the extractable quantitative data allow for adopting chemometric strategies for (i) method optimization and (ii) sample discrimination/classification in chemotaxonomic studies. In this regard, a targeted LC-MS/MS method for the determination of thirty-eight (poly)phenolic compounds in *Diospyros kaki* was optimized by Analytical Quality by Design (AQbD), a multivariate linear modelling approach. In detail, key chromatographic features like critical chromatographic resolution and analysis time, were used as model responses for the optimization of most relevant chromatographic parameters (e.g., column temperature, gradient time, pH etc.) [29]. The optimized method was applied to the accurate quantification of the selected phenolic compounds in commercially harvested persimmons, to assess

cultivar effect and postharvest treatments alterations by chemometrics analysis (i.e., univariate statistical inference and multivariate analysis) [28].

The development of modern LC-MS/MS instrumentation pushes towards analytical throughput, improving both chromatographic and mass spectrometric features of targeted platforms. In this context, the increase in analytical method performances is most requested by environmental chemistry, due to the growing demand of sensitivity and accuracy in the analysis of chemicals of emerging concern for environmental and human safety. The availability of larger volumes of injection and on-line sample handling in novel LC instrumentation answers this great demand. A large-volume direct injection (LVDI) method has been applied for the analysis of perfluoroalkyl substances (PFASs) in a relevant number of drinking, ground, and surface waters, as well as in influent and effluent wastewaters (TWWs) of sewage treatment facilities [24]. Thanks to the overall negligible ME (generally <|20%|), the LVDI targeted platform exhibited very good detection limits, in line with the most recent regulations on PFASs. Moreover, cluster and multivariate analysis highlighted the influence of sample origin on ME, and allowed for interpreting and predicting the signal suppression or enhancement as a function of physicochemical parameters of water samples. When ME is not negligible, on-line SPE-LC represents a "high-throughput" strategy, combining sample extraction and enrichment (i.e., sensitivity) with the clean-up for interference removal (i.e., ME). An example is given by the development of an on-line SPE-LC-MS/MS platform to the analysis of non-steroidal anti-inflammatory drugs in TWW sewage sludge extracted by Quick, Easy, Cheap, Effective, Rugged, and Safe (QuEChERS) method [21]. The developed method showed analytical sensitivities suitable for targeted analyte determination in real samples from tens of pg g^{-1} to low ng g^{-1} with a total analysis time of 30 min per sample. An on-line SPE-LC-MS/MS for the simultaneous analysis of twenty-three alkylphenols polyethoxylated and their derivatives, was developed exploiting a key feature of modern API sources, i.e. the polarity switching in ESI mode, as well as novel SPE polymeric sorbent phases [20]. The

combination of negative-to-positive single polarity switching with the on-line SPE resulted useful in the one-shot analysis of target analytes, reducing matrix-dependent source phenomena which are enhanced by continuous voltage variations. By adopting this targeted platform, very good sensitivities were obtained for most of the target analytes (0.0081–1.0 ng L^{-1}) in inlet and outlet wastewater samples, exhibiting also apparent R% in the 50-120% range.

1.4 LC-MS/MS platforms employed in this PhD thesis work

In this section, the wet and dry lab instruments, combined to structure the LC-MS/MS platforms of interest for this thesis work, will be described.
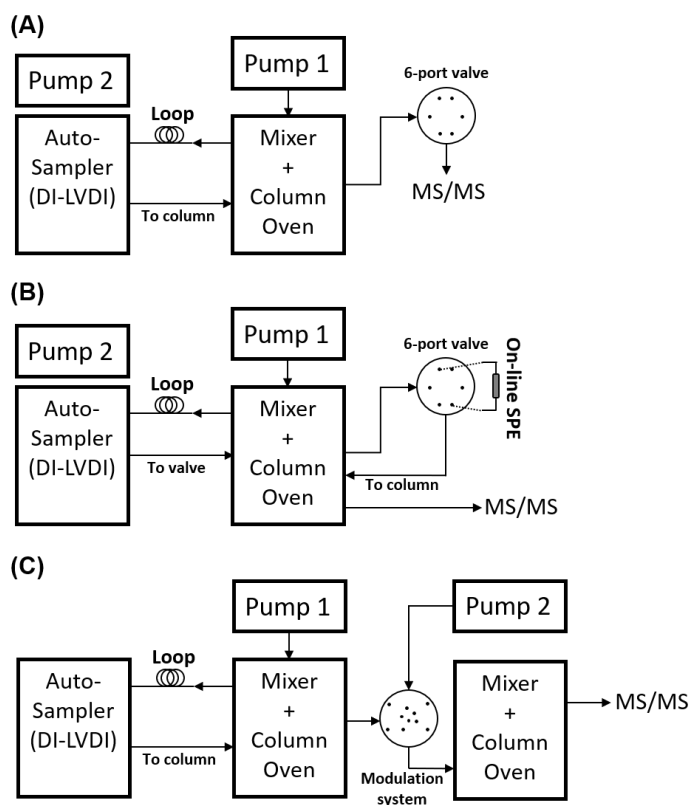
Figure 1.1 – Schemes of the LC configurations adopted in this thesis work for (A) large-volume (LVDI) or low-volume direct injection (DI), (B) on-line SPE sample pre-concentration and clean-up, and (C) 2D-LC or LC×LC analysis of complex mixtures.

1. General introduction

Figure 1.1 illustrates the general schemes of the LC systems for untargeted and targeted platforms, configured to perform (i) low- (DI) and larg-volume (LVDI) direct injection, (ii) on-line SPE pre-treatment, and (iii) 2D-LC analyses. All the displayed setups take advantage of valve modules used as sample diverter in the DI/LVDI configuration to preserve the analytical column, as sample manager of loading (pre-concentration and clean-up) and injection phases for on-line SPE, and as modulation system to transfer eluate fractions from the first to the second dimension in the 2D-LC/LC×LC configuration.
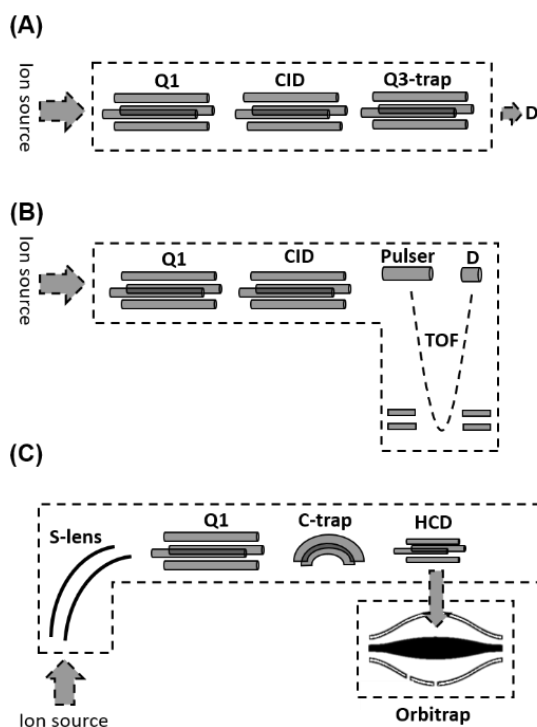
Figure 1.2 – MS analyzers adopted in this thesis work. (A) Hybrid QqQ used for targeted and accurate quantitative analysis, (B) Q-TOF and (C) LTQ/Q Exactive Plus Orbitrap instruments for HRMS untargeted/non-targeted analysis.

Figure 1.2 shows the MS (both low- and high-resolution) analyzers hyphenated with the LC systems. Targeted experiments are performed using a hybrid QqQ analyzer operating in MRM mode (10-25 ms of dwell time for each precursor/product transition) and equipped with an ESI probe capable of providing the polarity switch in hundreds of milliseconds. The untargeted studies

were conducted using two different high-resolution MS/MS analyzers, based on TOF (separation "in space") and Orbitrap (separation "in time") technologies. Both instruments are hybridized with upstream quadrupolar (or multipolar) systems acting as mass filters and/or collision cells to increase the resolving power of the downstream MS analyzer and to carry out MS/MS experiments. According to the application of intertest, these technologies were exploited to acquire the data in full scan mode, as well as in DIA or DDA fashion.

The raw data obtained from the LC-MS/MS platform has been processed to extract qualitative (untargeted) and quantitative (targeted) information through different dry lab strategies. In the first case, statistical (e.g., parametric, and non-parametric inference), descriptive (e.g., Kendrick mass defect), and computational (e.g., molecular networking) tools were used to decipher sample information and to carry out an accurate annotation of the feature (e.g., metabolites) of interest. The quantitative data were used both to evaluate the analytical throughput of the developed platforms (e.g, sensitivity, and accuracy), and for multivariate analysis strategies aimed at (i) classification/discrimination of the analyzed samples (e.g., class analysis and principal component analysis) or (ii) optimization of the analytical method itself (e.g., AQbD).

1.5   Outline of the thesis

The present work has been carried out to develop high throughput LC-MS/MS platforms exploiting both targeted and untargeted approaches capable to untangle the intricate data information from complex samples of environmental and nutritional interest. To this extent, different types of (analytically challenging) samples have been investigated, including fruit extracts, wastewaters, technical mixtures, and human biofluids (i.e., urine and serum).

The development and/or application of (i) LC-MS/MS wet lab methods and (ii) dry lab chemometric workflows will generate informative and accurate data that could help to increase the knowledge on the composition of the investigated samples. In this regard, when applicable, LC-MS/MS data have been publicly

1. General introduction

shared in available repositories (e.g., MassIVE) to contribute to dissemination within the Scientific Community. Brief descriptions of the researches presented in this PhD thesis is reported in the following lines.

In section 2 (article published on Journal of the Science of Food and Agriculture 101 (2021) 1239-1246), the quality of strawberry (*Fragaria* x *ananassa*, cultivar Camarosa) fruits irrigated with four types of TWWs was investigated, supporting the principles of environmental sustainability. The quantitative data of fruit yield, sucrose, fructose, glucose, total soluble (poly)phenols, total monomeric anthocyanins, and antiradical and antioxidant capacity were determined, together with those regarding a target group of most representative phenolic compounds of strawberry. Statistical and multivariate analysis evidenced a lower quality of fruits grown under TWW irrigation compared to those obtained under conventional cultivation conditions. However, the nutritional and nutraceutical parameters determined on fruits irrigated with TWWs were in line with data previously reported for strawberries purchased in the market or cultivated in research orchards, thus suggesting that the use of TWWs does not prevent the fruit marketability

In section 3 (article published on Antioxidants 10 (2021) 704), the profiling of the antioxidant and bioactive compounds occurring in *Rubus idaeus* L. and *Rubus occidentalis* L. raspberries was provided by the adoption of a comprehensive untargeted metabolomics approach developed with UHPLC-MS/MS, using the SWATH® acquisition protocol. The feature selection and annotation workflow, applied to the analysis of raspberry extracts in both polarities, allowed identifying 68 bioactive compounds mainly belonging to the classes of (poly)phenolic compounds, also providing first reports in raspberry fruits of ferulic acid glycosides and ellagitannin-like compounds. Chemometrics was used to highlight the features more related to the expression of a genotype effect within the *R. idaeus* species or between the two raspberry species herein investigated, also in correlation with the collected pomological data.

1. General introduction

In section 4 (article published on Journal of Chromatography A 1649 (2021) 462225), AQbD is used to optimize the chromatographic separation and mass spectrometric detection of a wide group of structurally heterogeneous model pharmaceutical compounds and their major transformation products. Octadecyl and pentafluorophenyl stationary phases, acetonitrile/methanol ratios and acidity of the eluents, column temperature, initial organic phase percentage, and elution gradient were investigated by AQbD, aiming at optimizing critical resolutions, sensitivities, and analysis time. The optimized method allowed for analysing the target analytes in a single chromatographic run, adopting a mixed acquisition mode based on scheduled acquisition windows comprising both single polarity and continuous polarity switching. For most investigated analytes the method provided high sensitivity, with detection limits in the sub-ng L^{-1} to low ng L^{-1} range, and for this reason it was applied to the direct injection analysis of effluents from four wastewater treatment plants and surface waters.

In section 5 (article published on Analytical and Bioanalytical Chemistry 414 (2022) 1841–1855), different chemometric selection workflows for potential exposure marker (PEM) discovery, using multivariate or univariate parametric or non-parametric data analyses, were comparatively tested and evaluated. The PEM selection protocols were applied to a small-sample-size untargeted LC-MS/MS study of a longitudinal set of serum samples from 20 volunteers after a single intake of (poly)phenolic-rich *Vaccinium myrtillus* and *Vaccinium corymbosum* supplements. According to study design and dataset structure, the non-parametric Games-Howell test was identified as the most suitable approach for the identification of significant features, thus minimizing the risk of false-positive retention. Among the forty-seven PEMs exhibiting a statistically significant postprandial kinetics, twelve were successfully annotated. In addition, the analysis of the area under the curve of the longitudinal dataset highlighted thirteen statistically significant PEMs discriminating the two interventions, including four intra-intervention relevant metabolites, as further confirmed by principal component analysis and sample classification.

1. General introduction

In section 6 (article published on Metabolites 12 (2022) 1005), the investigation of the metabolite discovery capacity and the possible extension of the annotation coverage of the Feature-Based Molecular Networking (FBMN) approach was performed in combination with public and contextual mass spectral libraries in comparison with the currently accepted non-commercial annotation protocols. The postprandial urinary metabolome analysis within a two-arm intervention on the intake of *Vaccinium myrtillus* and *Vaccinium corymbosum* supplements was selected as a case study. To further exploit the quantitative data obtained within the FBMN environment, the postprandial behaviour of the annotated metabolites was analyzed with Pearson product-moment correlation.

In section 7 (Analytica Chimica Acta 1232 (2022) 340485), an online comprehensive LC×LC-HRMS method was developed for the non-targeted identification of poly- and perfluorinated compounds (PFASs) in fire-fighting aqueous-film forming foams (AFFFs). The method exploited the combination of mixed-mode weak anion exchange-reversed phase with octadecyl stationary phase, separating PFASs according to ionic classes and chain length. Method development was carried out in two main steps, the first focused on the study of the orthogonality of different stationary phases, including mixed mode materials of interest, and the second focused on the two-dimensional method optimization of (i) elution conditions and (ii) technology assisted modulation strategies. The LC×LC-HRMS method applied to the characterization of AFFF mixtures, demonstrated the gain in terms of the number of detected features with respect to 1D-LC.

In section 8 (article currently in preparation), thermal desorption (TD) was investigated for the first time in on-line SPE-RP-LC-MS/MS, using water as SPE eluent. This desorption strategy was tested in the on-line SPE-LC-MS/MS analysis of a model group of structurally heterogeneous analytes including 34 gut (poly)phenolic metabolites (selected from the annotated compounds in the untargeted studies) characterized by a wide range of physicochemical and structural properties, including isobaric compounds, thus representing a

1. General introduction

challenging analytical issue. At first, the chromatographic separation of the model analytes was investigated using different RP stationary phases and eluent combinations, focusing on the resolution of critical isobaric groups. Then, TD was studied in comparison with organic desorption followed by post-column dilution and with conventional desorption demonstrating the higher performance of this novel desorption approach in terms of analyte focusing and sensitivity. Finally, the on-line TD-SPE-RP-LC-MS/MS method was evaluated for matrix effect and recovery in urine, as relevant applications in nutrimetabolomics.

References

[1] B.L. Milman, I.K. Zhurkovich, The chemical space for non-target analysis, TrAC Trends in Analytical Chemistry, 97 (2017) 179-187 https://doi.org/10.1016/j.trac.2017.09.013.

[2] K. Broeckhoven, G. Desmet, Advances and innovations in liquid chromatography stationary phase supports, Analytical Chemistry, 93 (2020) 257-272 https://doi.org/10.1021/acs.analchem.0c04466.

[3] J. De Vos, K. Broeckhoven, S. Eeltink, Advances in ultrahigh-pressure liquid chromatography technology and system design, Analytical chemistry, 88 (2016) 262-278 https://doi.org/10.1021/acs.analchem.5b04381.

[4] L. Wang, W. Wei, Z. Xia, X. Jie, Z.Z. Xia, Recent advances in materials for stationary phases of mixed-mode high-performance liquid chromatography, TrAC Trends in Analytical Chemistry, 80 (2016) 495-506 https://doi.org/10.1016/j.trac.2016.04.001.

[5] B.W. Pirok, A.F. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, Journal of separation science, 41 (2018) 68-98 https://doi.org/10.1002/jssc.201700863.

[6] A. Stavrianidi, A classification of liquid chromatography mass spectrometry techniques for evaluation of chemical composition and quality control of traditional medicines, Journal of chromatography A, 1609 (2020) 460501 https://doi.org/10.1016/j.chroma.2019.460501.

[7] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC Trends in Analytical Chemistry, 82 (2016) 425-442 https://doi.org/10.1016/j.trac.2016.07.004.

[8] M. Roca, M.I. Alcoriza, J.C. Garcia-Cañaveras, A. Lahoz, Reviewing the metabolome coverage provided by LC-MS: Focus on sample preparation and chromatography-A tutorial, Analytica Chimica Acta, 1147 (2021) 38-55 https://doi.org/10.1016/j.aca.2020.12.025.

[9] K. Ortmayr, T.J. Causon, S. Hann, G. Koellensperger, Increasing selectivity and coverage in LC-MS based metabolome analysis, TrAC Trends in Analytical Chemistry, 82 (2016) 358-366 https://doi.org/10.1016/j.trac.2016.06.011.

[10] J. Guo, T. Huan, Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics, Analytical Chemistry, 92 (2020) 8072-8080 https://doi.org/10.1021/acs.analchem.9b05135.

[11] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L. Brennan, C. Brunius, A. Bub, F. Capozzi, Nutrimetabolomics: an integrative action for metabolomic analyses in human nutritional studies, Molecular nutrition & food research, 63 (2019) 1800384 https://doi.org/10.1002/mnfr.201800384.

[12] L. Renai, C. Ancillotti, M. Ulaszewska, M. Garcia-Aloy, F. Mattivi, R. Bartoletti, M. Del Bubba, Comparison of chemometric strategies for potential exposure marker discovery and false-positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of Vaccinium fruit supplements, Analytical and Bioanalytical Chemistry, 414 (2022) 1841-1855 https://doi.org/10.1007/s00216-021-03815-5.

[13] F. Menger, P. Gago-Ferrero, K. Wiberg, L. Ahrens, Wide-scope screening of polar contaminants of concern in water: A critical review of liquid chromatography-high resolution mass spectrometry-based strategies, Trends in Environmental Analytical Chemistry, 28 (2020) e00102 https://doi.org/10.1016/j.teac.2020.e00102.

[14] E. Rampler, D. Egger, H. Schoeny, M. Rusz, M.P. Pacheco, G. Marino, C. Kasper, T. Naegele, G. Koellensperger, The power of LC-MS based multiomics: Exploring adipogenic differentiation of human mesenchymal stem/stromal cells, Molecules, 24 (2019) 3615 https://doi.org/10.3390/molecules24193615.

[15] W. Lv, X. Shi, S. Wang, G. Xu, Multidimensional liquid chromatography-mass spectrometry for metabolomic and lipidomic analyses, TrAC Trends in Analytical Chemistry, 120 (2019) 115302 https://doi.org/10.1016/j.trac.2018.11.001.

[16] N. Rauniyar, Parallel reaction monitoring: a targeted experiment performed using high resolution and high mass accuracy mass spectrometry, International journal of molecular sciences, 16 (2015) 28566-28581 https://doi.org/10.3390/ijms161226120.

[17] J.G. van der Gugten, Tandem mass spectrometry in the clinical laboratory: a tutorial overview, Clinical Mass Spectrometry, 15 (2020) 36-43 https://doi.org/10.1016/j.clinms.2019.09.002.

[18] A. Kruve, R. Rebane, K. Kipper, M.-L. Oldekop, H. Evard, K. Herodes, P. Ravio, I. Leito, Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part II, Analytica chimica acta, 870 (2015) 8-28 https://doi.org/10.1016/j.aca.2015.02.016.

[19] L. Renai, C.V.A. Scordo, A. El Ghadraoui, S. Santana-Viera, J.J.S. Rodriguez, S. Orlandini, S. Furlanetto, D. Fibbi, D. Lambropoulou, M. Del Bubba, Quality by design optimization of a liquid chromatographic-tandem mass spectrometric method for the simultaneous analysis of structurally heterogeneous pharmaceutical compounds and its application to the rapid screening in wastewater and surface water samples by large volume direct injection, Journal of

1. General introduction

Chromatography A, 1649 (2021) 462225 https://doi.org/10.1016/j.chroma.2021.462225.

[20] L. Ciofi, C. Ancillotti, U. Chiuminatto, D. Fibbi, B. Pasquini, M.C. Bruzzoniti, L. Rivoira, M. Del Bubba, Fully automated on-line solid phase extraction coupled to liquid chromatography–tandem mass spectrometry for the simultaneous analysis of alkylphenol polyethoxylates and their carboxylic and phenolic metabolites in wastewater samples, Analytical and bioanalytical chemistry, 408 (2016) 3331-3347 https://doi.org/10.1007/s00216-016-9403-5.

[21] D. Rossini, L. Ciofi, C. Ancillotti, L. Checchini, M.C. Bruzzoniti, L. Rivoira, D. Fibbi, S. Orlandini, M. Del Bubba, Innovative combination of QuEChERS extraction with on-line solid-phase extract purification and pre-concentration, followed by liquid chromatography-tandem mass spectrometry for the determination of non-steroidal anti-inflammatory drugs and their metabolites in sewage sludge, Analytica chimica acta, 935 (2016) 269-281 https://doi.org/10.1016/j.aca.2016.06.023.

[22] T. Tome, N. Žigart, Z. Časar, A. Obreza, Development and optimization of liquid chromatography analytical methods by using AQbD principles: Overview and recent advances, Organic Process Research & Development, 23 (2019) 1784-1802 https://doi.org/10.1021/acs.oprd.9b00238.

[23] C. Ancillotti, L. Ciofi, D. Pucci, E. Sagona, E. Giordani, S. Biricolti, M. Gori, W.A. Petrucci, F. Giardi, R. Bartoletti, Polyphenolic profiles and antioxidant and antiradical activity of Italian berries from Vaccinium myrtillus L. and Vaccinium uliginosum L. subsp. gaultherioides (Bigelow) SB Young, Food Chemistry, 204 (2016) 176-184 https://doi.org/10.1016/j.foodchem.2016.02.106.

[24] L. Ciofi, L. Renai, D. Rossini, C. Ancillotti, A. Falai, D. Fibbi, M.C. Bruzzoniti, J.J. Santana-Rodriguez, S. Orlandini, M. Del Bubba, Applicability of the direct injection liquid chromatographic tandem mass spectrometric analytical approach to the sub-ng L− 1 determination of perfluoro-alkyl acids in waste, surface, ground and drinking water samples, Talanta, 176 (2018) 412-421 https://doi.org/10.1016/j.talanta.2017.08.052.

[25] C.V.A. Scordo, L. Checchini, L. Renai, S. Orlandini, M.C. Bruzzoniti, D. Fibbi, L. Mandi, N. Ouazzani, M. Del Bubba, Optimization and validation of a method based on QuEChERS extraction and liquid chromatographic–tandem mass spectrometric analysis for the determination of perfluoroalkyl acids in strawberry and olive fruits, as model crops with different matrix characteristics, Journal of Chromatography A, 1621 (2020) 461038 https://doi.org/10.1016/j.chroma.2020.461038.

[26] C. Seger, L. Salzmann, After another decade: LC–MS/MS became routine in clinical diagnostics, Clinical biochemistry, 82 (2020) 2-11 https://doi.org/10.1016/j.clinbiochem.2020.03.004.

[27] C. Ancillotti, L. Ciofi, D. Rossini, U. Chiuminatto, J. Stahl-Zeng, S. Orlandini, S. Furlanetto, M. Del Bubba, Liquid chromatographic/electrospray ionization quadrupole/time of flight tandem mass spectrometric study of polyphenolic composition of different Vaccinium berry species and their

comparative evaluation, Analytical and bioanalytical chemistry, 409 (2017) 1347-1368 https://doi.org/10.1007/s00216-016-0067-y.

[28] C. Ancillotti, M. Ulaszewska, F. Mattivi, M. Del Bubba, Untargeted metabolomics analytical strategy based on liquid chromatography/electrospray ionization linear ion trap quadrupole/orbitrap mass spectrometry for discovering new polyphenol metabolites in human biofluids after acute ingestion of Vaccinium myrtillus berry supplement, Journal of the American Society for Mass Spectrometry, 30 (2019) 381-402 https://doi.org/10.1007/s13361-018-2111-y.

[29] C. Ancillotti, S. Orlandini, L. Ciofi, B. Pasquini, C. Caprini, C. Droandi, S. Furlanetto, M. Del Bubba, Quality by design compliant strategy for the development of a liquid chromatography–tandem mass spectrometry method for the determination of selected polyphenols in Diospyros kaki, Journal of Chromatography A, 1569 (2018) 79-90 https://doi.org/10.1016/j.chroma.2018.07.046.

2 Productivity and nutritional and nutraceutical value of strawberry fruits (*Fragaria* x *ananassa* Duch.) cultivated under irrigation with treated wastewaters.

Supplementary materials:

https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fjsfa.10737&file=jsfa10737-sup-0001-Supinfo.docx

Section 2

Abstract

Agriculture represents a productive sector typically characterized by a high water demand, whilst FW availability is a problem of increasing concern in the world and FW resources are becoming insufficient for sustaining agricultural irrigation. The reuse of treated wastewaters (TWWs) for crop irrigation could be an efficient tool of reducing water shortage. Hence, this study evaluated the food quality of *Fragaria* x *ananassa* (cultivar Camarosa) fruits irrigated with four kinds of treated wastewaters (TWWs). Strawberries were analysed for yield, sucrose, fructose, glucose, total soluble polyphenols (TSP), total monomeric anthocyanins (TMA), as well as antiradical and antioxidant capacity. In addition, a targeted quantification of the most representative phenolic compounds of strawberry was performed. TWWs complied the Italian ministerial decree 185/2003 for wastewater reuse with very few exceptions, mainly represented by chloride concentrations (258-643 mg/L vs a legal threshold of 250 mg/L). The reuse of TWWs reduced fruit yield (10-26%) compared to irrigation with tap water as control. Irrigation with TWWs gave also rise to the decrease of total sugars (14-26%), TSP (2-10%) and TMA (29-49%). Individual phenolic acids, flavonols and flavanols were quite stable in response to the irrigation with TWWs, whereas anthocyanidins significantly decreased. Although TWWs negatively affected fruit quality, nutritional and nutraceutical parameters herein determined were in line with data previously reported for strawberries purchased in the market or cultivated in research orchards, thus suggesting that the use of TWWs does not prevent the fruit marketability.

Keywords: fruit yield; sugars; polyphenols; wastewater reuse; circular economy

2.1 Introduction

Agriculture represents a productive sector typically characterized by a high water demand. According to the European Environment Agency, a third of the water use in Europe goes to the agricultural sector, most of it for crop irrigation [1], and, as recently pointed out by the United Nations World Water Assessment Program [2],

about 70% of worldwide freshwater (FW) withdrawals is used for agricultural irrigation. On the other hand, limited FW availability is a problem of increasing concern in the world and FW resources are becoming insufficient to efficiently sustain agricultural irrigation, mainly due to climate-related conditions. In fact, water scarcity is in most cases a climate-bound regional problem and affects many areas of the Earth's planet, including Middle East, North Africa [3], but also Southern Europe, including Italy [4]. The reuse of non-conventional waters for irrigation, such as treated wastewaters (TWWs) of municipal or mixed municipal/industrial origin, could be an efficient tool of reducing water shortage [5], reason why TWWs reuse is becoming a widely adopted practice in agriculture [2]. Moreover, soils and plants can benefit from the fertilizing effect of wastewater [6]. However, TWWs may contain chemical and bacteriological contaminations that can affect crop safety. For this reason, many countries have developed their own regulations in the field of water reuse [7]. For example, in Italy, wastewaters are allowed to be reused for the irrigation of crops intended for both human and animal consumption, whether a number of chemical and biological properties meet the limits established by a specific regulation on wastewater reuse [8]. Moreover, TWWs often exhibit physicochemical and/or chemical properties (e.g. pH, conductivity, sodium and chloride ions), which may negatively affect crop productivity and/or quality [9]. Quality in food is a combination of different attributes (e.g. sugars, minerals and bioactive compounds), which affect organoleptic properties, as well as nutritional and nutraceutical values. These compounds are susceptible to significant variations, depending on climate conditions and agronomic practices.[10] Generally, the quality of vegetables and fruits irrigated with TWWs has been commonly evaluated through their main pomological parameters related to product marketability, reporting slight differences compared to traditional watering techniques.[11,12] Conversely, the impact of crop irrigation by TWWs on nutritional and nutraceutical value is poorly described in literature. More in detail, irrigation with TWWs of short-term crops, like strawberries, does not seem to promote significant variations of the principal

nutritional and nutraceutical values [13], whereas on long-term crops, like olive trees, the effect of an extended TWW irrigation increased the level of β-carotene and total tocopherols of olive oil [3]. Among crop species that can be investigated for their quality in response to irrigation with TWWs, strawberry (Fragaria x ananassa Duch.) is certainly a very attractive fruit due to its unique organoleptic characteristics, as well as overall fruit nutritional and nutraceutical attributes [14], reasons why strawberry is widely appreciated by consumers. In fact, strawberry covers an important place in the horticultural industry, particularly in the Mediterranean countries [15], which produce around 1.6 million tons annually, almost 18% of the world production [5]. However, these countries are notoriously suffering from limited water resources, which clash with the high demand for water to irrigate strawberry.

Based on the aforementioned considerations, in this study, strawberry plants were grown under irrigation with four types of TWWs, characterized by different physicochemical attributes (e.g. different level of salinity), using tap water (TW) as control. More in detail, Camarosa cultivar was selected due to its lower salt tolerance threshold, compared to other varieties [16]. Strawberry quality was evaluated through the analysis of sucrose, glucose, and fructose as essential nutritional parameters.[17] Total soluble polyphenols (TSP), total monomeric anthocyanins (TMA), as well as radical scavenging and antioxidant activities (RSA and AA) were also analysed, as important nutraceutical attributes.[18] Moreover, some phenolic compounds previously highlighted as important constituents of the phenolic fraction of Fragaria fruits[19-21] were determined to further characterize fruit nutraceutical quality under non-conventional irrigation practices. Through this experimental design, the following hypotheses will be verified: (i) the use of different TWWs and TW impart significant differences in the nutritional and/or nutraceutical quality of the strawberries obtained; (ii) the nutritional and nutraceutical quality of the fruits obtained by non-conventional irrigation is high enough to allow their marketability.

2.2 Materials and methods

Standards, reagents, solvents and materials used in this study are described in section S1 of the *Supplementary materials*.

2.2.1 Sample origin

Young fridge stored certified *Fragaria* x *ananassa* plants (Camarosa cultivar) were grown outdoor from March to July 2017 (see Section S2 of the *Supplementary materials* for details). Plants were irrigated with four TWWs collected in wastewater treatment plants (WWTPs) managed by GIDA S.p.A. (Prato, Italy). More in detail, the TWWs derived from the following WWTPs: (i) "Baciacavallo" (TWW1), (ii) "Macrolotto 1" (TWW2), (iii) "Macrolotto 2" (TWW3), and (iv) "Calice" (TWW4). TW was used as control. WWTPs description is reported in the Section S3 of the Supplementary materials. Physicochemical, chemical, and microbiological parameters reported in the Italian regulation on wastewater reuse [8] were determined in TWWs and the results are shown in **Table S1** of the *Supplementary materials*, together with data regarding TW, which were taken from a public database [22]. Strawberry fruits were harvested when characterized by a red colour all over the fruit. The collected strawberries were transported to the laboratory, gently washed with distilled water, dried with paper towel and finally weighted in order to determine the fruit yield. All fruits from each plant were separately freeze-dried and stored at -20 °C until analysis.

2.2.2 Extraction of sugars and phenolic compounds

Sugars and phenolic compounds were extracted by the same procedure [23], using raffinose, myricetin and petunidin-3-O-arabinoside for the evaluation of the apparent recovery [24]. Full details of the extraction procedure are reported in the Section S4 of the *Supplementary materials*.

2.2.3 Analysis of sugars

Fructose, glucose and sucrose were instrumentally determined by liquid chromatography (LC), coupled with evaporative light scattering detection (ELSD) after frontal elution of the extracts on Supelclean LC-18 SPE Tubes. Individual and total sugars were expressed as mg g^{-1} d.w. and mmol g^{-1} d.w., respectively. Full details of the LC-ELSD analysis are reported in the Section S5 of the *Supplementary materials*, whereas figures of merit of the method are shown in **Table S2**.

2.2.4 Analysis of TSP, TMA, RSA and AA

TSP, TMA, RSA and AA were determined on the extracts using spectrophotometric methods. TSP were analysed according to the Folin-Ciocalteu method [23], using calibration lines prepared with (+)-catechin (see Section S6 of the *Supplementary material* for full details). TMA were determined with the pH differential method [25] using pelargonidin-3-glucoside as reference standard. RSA was determined through the methods based on 2,2-diphenyl-1-picrylhydrazyl (DPPH) and 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS) radicals.[26,27] AA was measured through the Ferric Reducing Antioxidant Power (FRAP) assays [28]. Results were expressed as micromoles of Trolox equivalents per gram of fruit on a dry weight basis (µmol of Trolox g^{-1} d.w.). TSP and TMA were used for the evaluation of the relative recovery percentage of sequential extractions (see Sections S7 of the *Supplementary material*).

2.2.5 Analysis of individual phenolic compounds

Selected individual phenolic acids, chalcones, flavanols, flavonols and anthocyanins were analysed by LC hyphenated with electrospray ionization (ESI) triple quadrupole tandem mass spectrometry (MS/MS), using a Shimadzu (Kyoto, Japan) chromatographic system coupled with a 5500 QTrapTM mass spectrometer (Sciex, Ontario, Canada). Full details of LC-MS/MS analysis of

targeted phenolic compounds are reported in the Section S8 of the *Supplementary material*.

2.2.6   *Statistical analysis*

The analysis of variance, the non-parametric Games-Howell test for multiple comparison of the mean concentration values and the Pearson correlation test were performed by using Minitab®17.1.0 (Minitab Inc., State College, PA, USA). Principal component analysis (PCA) and cluster analysis (CA) were performed using Minitab®17.1.0. Quality control of PCA was carried out on the mix of extracts of strawberry fruits grown under irrigation with the four investigated TWWs and TW as control (QCs), verifying if object scores are close to the origin of new coordinates in the principal component (PC) plot.

2.3   **Results and discussion**

2.3.1   *TWWs characterization*

Table S1 of the *Supplementary materials* illustrates the physicochemical, chemical and biological parameters of the TWWs used in this study, which are foreseen by the Italian Ministerial Decree 185/2003 [8], regulating the wastewater reuse for various applications, including the agricultural irrigation. **Table S1** includes also the values of these parameters determined in TW, as well as the limits reported in the M.D. 185/2003. The values determined in TWWs complied the thresholds with some exceptions. More in detail, among physicochemical and chemical parameters, TSS slightly exceeded the limit established by the M.D. 185/2003 (i.e. 10 mg L^{-1}) for TWW1 (11±4 mg L^{-1}) and TWW4 (12±4 mg L^{-1}), and ammonia was found just above the legal threshold (i.e. 2 mg L^{-1}) in TWW1 (2.1±1.3 mg L^{-1}). Exceedances of the M.D. 185/2003 limits were much more accentuated and generalized (i.e. in all the TWWs investigated) for the chloride ion, the mean concentrations of which were in the range of 258-643 mg L^{-1} (legal threshold 250 mg L^{-1}). Sodium adsorption ratio (SAR), although within the limit established by the M.D. 185/2003 (i.e. 10), was also a critical parameter for TWW reuse in agriculture, being it included from 9.0 (for TWW3) to 9.5 (for all the other

TWWs). Conductivity was a further parameter worth to be mentioned, since high values were observed in TWWs (1322-2428 μS cm^{-1}), compared to those considered suitable for crop irrigation [29]. Overall, the remarkable concentrations of chloride ion, together with the high values of SAR and conductivity, highlight potential problems in the use of TWWs for irrigation purposes. However, it should be noted that these waters represent an important source of nutrients for plant growth, since they respectively contain 5-8 mg L^{-1} of nitrogen and 300-900 μg L^{-1} of phosphorus.

2.3.2 *Fruit yield*

Table 1 illustrates the fruit yield obtained with the four TWWs and TW (control). The irrigation with TWWs influenced fruit yield, resulting in a general decrease of productivity. The trend of fruit yield followed the order TW>TWW4≈TWW3=TWW2>TWW1. More in detail, plants irrigated with TWWs showed a reduced fruit production compared to control from 10% with TWW4 to 26% with TWW1, the latter exhibiting by far the highest level of salinity, as measured by electrical conductivity (2428 μS cm^{-1} in TWW1 vs 1322-1647 μS cm^{-1} in the other TWWs and 872 μS cm^{-1} in TW). Interestingly, a similar reduction in productivity (12-24% depending on the kind of irrigation system) was previously reported for Camarosa strawberry [13] irrigated with a TWW, which displayed a conductivity comparable to TWWs used in this study. A stronger yield reduction (38-63%) was highlighted in various *Fragaria* x *ananassa* varieties in response to increasing conductivity levels (from 700 to 2500 μS cm^{-1}) of irrigation water [30]. These findings evidenced the presence of a stress condition in strawberry plants irrigated with TWWs, in agreement with the aforementioned higher levels of chloride, SAR and conductivity in TWWs than in TW (**Table S1**). Salinity may compromise the plant water ability absorption, since ions in soil solution force plant to further lower its water potential to maintain a proper water supply from soil [31], causing a plant water-deficit condition, which inhibited plant growth and productivity [32]. Moreover, a specific toxicity of chloride ion may contribute to the yield reduction observed in this study, since chloride

concentrations as high as 150 mg L^{-1} exhibited toxicity towards Camarosa strawberries with significant effects on fruit production [16]. The yield could be further compromised if long-term irrigation with saline TWWs is carried out. However, it should be considered that the soilless cultivation of horticultural products, including strawberry, usually involves plants and substrate replacement every two vegetative cycles (i.e. one-two years) owing to the decrease of the production performances observed after this period [33,34], making therefore less critical the impact of irrigation with TWWs.

Table 1 – Mean ± SD (n = 63–66, depending on the treatment) values of the fruit yield of strawberry plants irrigated with TWWs and TW as control.

Treatment	g f.w./plant
TW	89 (10) a
TWW1	66 (7) b
TWW2	74 (9) c
TWW3	73 (9) c
TWW4	80 (10) c

The yield is expressed as grams of fruit fresh weight per plant (g FW plant–1). Values with different lowercase letters are statistically different according to the Games–Howell multicomparison test (P < 0.05).

2.3.3   Sugars

Table 2 shows the concentrations of fructose, glucose and sucrose determined in strawberry fruits obtained under irrigation with TW and TWWs. Sugar levels found herein were in line with the range elsewhere reported for Camarosa fruits purchased in the market or cultivated in soilless systems in research orchards [35-38], thus demonstrating that, from this viewpoint, the fruit quality is high enough to guarantee their marketability. However, significant variations (p<0.05) were observed among treatments. In fact, fruits irrigated with TWWs showed significantly lower values of individual and total sugar compared with control fruits. Fruits grown under irrigation with TWW2 and TWW3 showed comparable concentrations of total sugars, fructose and glucose. The greater influence on sugar concentrations of the irrigation by TWWs was highlighted for strawberries obtained with TWW1 and TWW4, which displayed the lowest individual and total sugar values (i.e. 1.98 and 2.06 mmol g^{-1} d.w., respectively), approximately 25%

lower than control fruits. The lower abundance of individual sugars in strawberry fruits might be ascribable to the salinity of the TWWs used for irrigation, which showed Cl- concentration exceeding Italian legal limits for wastewater reuse (**Table S1**). In fact, the high Cl- concentrations could have caused a water deficit in the strawberry plants. The reduction of carbohydrates was probably linked to the consumption of photoassimilates for osmotic adjustment, as previously reported for fruits of strawberry plants cultivated in soils characterized by high NaCl contents [39,40].

2.3.4   *TSP, TMA, RSA and AA*

Mean values of TSP, TMA, DPPH-RSA, ABTS-RSA, and FRAP-AA determined in strawberry fruits in response to the irrigation with TW and TWWs are shown in **Table 2**. The treatments exhibited quite similar TSP concentrations, being the highest variation (about 10%) observed between TWW1 (2521 mg catechin 100 g^{-1} d.w.) and control (2807 mg catechin 100 g^{-1} d.w.). This trend was also found elsewhere on Camarosa strawberries irrigated with a tertiary TWW characterized by conductivity and concentrations of BOD5, COD, Ntot and Ptot similar to those of TWWs tested in this study [13]. The comparison between TWW1 (the most salty TWW) and TW was the only one providing a statistically significant difference ($p<0.05$). The antiradical and antioxidant activity parameters behaved in a very similar way to the TSP, showing a significant linear correlation each other ($r=0.923$-0.988, $p<0.05$) and with TSP itself ($r=0.903$-0.960, $p<0.05$). In contrast to findings obtained for TSP, DPPH, ABTS, and FRAP, irrigation with wastewater significantly affected TMA values, as total anthocyanins in control fruits (610 mg pelargonidin-3-O-glucoside 100 g^{-1} d.w.) were up to twofold higher than those found in fruits treated with TWWs (310-437 mg pelargonidin-3-O-glucoside 100 g^{-1} d.w.). More in detail, the irrigation with TWWs gave rise in all cases to statistically significant decreases of this parameter compared to control. It is however remarkable that TSP and TMA concentrations of strawberries produced with TWWs were included in the range of values reported in literature for Camarosa fruits purchased in the market or produced in research

orchards [1,13,41,42]. Therefore, fruits irrigated with TWWs demonstrated a nutraceutical quality in line with their marketability.

Table 2 – Mean ± SD (n = 3) values of individual (mg g^{-1} DW) and total sugars (mmol g^{-1} DW), total soluble polyphenols (TSP, mg catechin 100 g^{-1} DW), total monomeric anthocyanins (TMA, mg pelargonidin-3-O-glucoside 100 g^{-1} DW), antiradical and antioxidant activities, as measured by DPPH, ABTS and FRAP methods (μmol Trolox g^{-1} DW) in strawberry plants irrigated with TW and TWWs.

Parameters	TW	TWW$_1$	TWW$_2$	TWW$_3$	TWW$_4$
Sugars					
Fructose	207 (2) a	146 (2) b	162 (5) c	167 (6) c	153 (3) d
Glucose	182 (2) a	132 (2) b	148 (5) cd	155 (2) c	140 (4) d
Sucrose	184 (2) a	156 (1) b	174 (3) c	191 (5) a	157 (9) b
Total sugars	2.67 (0.03) a	1.98 (0.02) b	2.20 (0.08) d	2.32 (0.05) d	2.06 (0.07) b
Phenolic compounds					
TSP	2807 (151) a	2521 (170) b	2680 (173) ab	2683 (71) a	2770 (149) a
TMA	610 (43) a	399 (62) bc	437 (26) b	393 (56) bc	310 (30) c
Antioxidant/antiradical activities					
DPPH	311 (25) a	263 (15) b	272 (19) ab	278 (23) ab	311 (8) a
ABTS	376 (46) a	316 (29) a	333 (29) a	345 (21) a	365 (42) a
FRAP	426 (46) a	355 (9) b	372 (11) b	396 (11) ab	408 (23) ab

The yield is expressed as grams of fruit fresh weight per plant (g FW plant–1). Values with different lowercase letters are statistically different according to the Games–Howell multicomparison test (P < 0.05).

2.3.5   *Individual phenolic compounds*

Table 3 shows the concentrations of targeted phenolic compounds (i.e. principal phenolic acids, chalcones, flavanols, flavonols and anthocyanins) herein used as further indicators of the quality of strawberry fruits obtained by irrigation with TW and TWWs. Table 3 also provides abbreviations of targeted analytes, which are used below. In the whole set of treatments, the majority of target analytes showed a signal-to-noise ratio higher than 10, being therefore successfully quantified. CHL, QUE-GAL, and CYA-GAL were determined only in fruits produced with TWW3, although at very low concentrations (≤0.31 mg 100 g^{-1} d.w.). Moreover, CAF, QUE, QUE-RHA and PHL were never quantified in the investigated samples. Similar patterns of relative abundance were highlighted for targeted phenolic compounds, irrespective of the use of TWWs or TW for irrigation. More in detail, in all samples, PEL-GLU was by far the most abundant compound (161-343 mg 100g^{-1} d.w.), accounting for 74-84% and 47-67% of total

individual anthocyanins and total individual phenolic compounds, respectively. Other predominant compounds were PB2 (29-54 mg 100 g^{-1} d.w.), CAT (40-49 mg 100 g^{-1} d.w.), EA (15-26 mg 100 g^{-1} d.w.), CYA (11-17 mg 100 g^{-1} d.w.), CYA-GLU (9-12 mg 100 g^{-1} d.w.), and PEL-RUT (20-40 mg 100g^{-1} d.w.). Literature data related to Camarosa strawberries obtained in soilless systems using fresh water for irrigation, confirmed this trend [1,43]. A general concentration increase was evidenced for non-anthocyanin phenolic compounds in fruits treated with TWWs compared to those irrigated with TW (**Table 3**). These differences were statistically significant only in few cases, such as PB1 for TWW1 and TWW4, and PB2 for TWW2 and TWW4. However, when the total concentration of these compounds was considered, the increase was remarkable (percentage increase of 15-29%) and statistically significant in all cases. Conversely, individual anthocyanins evidenced a concentration decrease in response to the use of TWWs in almost all cases. In particular, PEL-GLU and PEL-RUT were significantly lower in fruits produced with TWW1 and TWW4 compared to TW, whereas the use of TWW2 and TWW3 did not provide statistically significant reductions in concentration. A slight concentration decrease was observed for CYA-GLU in response to the use of all TWWs, but the differences were not statistically significant. An opposite behaviour was found for CYA, which was more abundant in fruits produced under irrigation with TWWs. Total concentration of the quantified anthocyanins followed the trend of the predominant individual anthocyanins (i.e. the two pelargonidins), being it statistically lower in fruits irrigated with TWW1 and TWW4, compared to TW. Interestingly, the sum of the concentrations of targeted individual anthocyanins represented a significant percentage (about 70-90%, depending on the sample considered) of total anthocyanins spectrophotometrically determined as TMA (see **Table 1**). Hence, the group of individual anthocyanins herein selected seems to give a representative picture of the whole set of anthocyanins occurring in strawberry fruits. In this regard, it should be noted that total concentrations of

individual anthocyanins showed some correlation with TMA values (r=0.795, p=0.108).

2.3.6 *Multivariate analysis*

In order to summarize the set of information obtained from the analysis of phenolic compounds in the 18 strawberry samples (including QCs), and to highlight more easily the effects of the irrigation with TWWs and TW, a multivariate elaboration of the autoscaled original data was performed by means of PCA and CA. These data elaborations included the 19 phenolic compounds quantified in at least one strawberry sample. As shown in Table S5 (Section S9 of the *Supplementary material*) four principal components (PCs), characterized by eigenvalues > 1 and accounting for percentages of explained variances (EV%) of 38.7%, 20.4%, 15.3% and 10.0%, were obtained (total EV%=84.4). However, the contributions of each variable to the four significant PCs were not well differentiated, since only few variables evidenced remarkable differences among the four components in terms of absolute values of loadings. More in detail, the highest differences among loadings within a same PC have been highlighted for (i) QUE-GLU, PEO-GLU and PEL-RUT in PC1, (ii) KAM-RUT and especially PHL-GLU in PC2, (iii) FER, CAT, PB2, and CYA in PC3, (iv) KAM-GLU and CYA-GLU in PC4 (see **Table S5**).

Table 3 – Mean ± SD (n = 3) values of selected phenolic acids, chalcones, flavanols, flavonols and anthocyanins (mg/100 g DW) in strawberry irrigated with TW and TWWs.

Compounds	Abbreviation	TW	TWW$_1$	TWW$_2$	TWW$_3$	TWW$_4$
ESI (-)						
Chlorogenic acid	CHL	0.04*-0.08**	0.04*-0.08**	0.04*-0.08**	0.31 (0.03)	0.04*-0.08**
Caffeic acid	CAF	<0.03*	<0.03*	<0.03*	<0.03*	<0.03*
Ferulic acid	FER	3.9 (0.3) ab	4.4 (0.3) ab	5.3 (0.6) ab	5.5 (0.5) a	3.6 (0.4) b
Ellagic acid	EA	24 (2) ab	26 (2) a	15 (1) b	22 (2) ab	26 (3) ab
Quercetin	QUE	<0.004*	<0.004*	<0.004*	<0.004*	<0.004*
Quercetin-3-*O*-galactoside	QUE-GAL	<0.02*	<0.02*	<0.02*	0.07 (0.01)	<0.02*
Quercetin-3-*O*-glucoside	QUE-GLU	1.3 (0.1) a	1.8 (0.2) ab	1.4 (0.1) a	1.2 (0.2) a	1.8 (0.1) b
Quercetin-3-*O*-rutinoside	QUE-RUT	0.79 (0.09) a	1.0 (0.1) a	0.82 (0.07) a	0.69 (0.04) a	0.9 (0.1) a
Quercetin-3-*O*-rhamnoside	QUE-RHA	<0.02*	<0.02*	<0.02*	<0.02*	<0.02*
Kaempferol-3-*O*-glucoside	KAM-GLU	1.9 (0.3) a	2.2 (0.2) a	1.5 (0.1) a	1.7 (0.2) a	2.0 (0.2) a
Kaempferol-3-*O*-rutinoside	KAM-RUT	1.39 (0.09) a	1.4 (0.1) a	1.3 (0.1) a	1.1 (0.1) ab	0.75 (0.08) b
Procyanidin A2	PA2	<0.11*	0.11*-0.26**	<0.11*	0.11*-0.26**	<0.11*
Procyanidin B1	PB1	4.5 (0.3) a	6.1 (0.5) b	5.4 (0.4) ab	4.8 (0.8) ab	5.9 (0.4) b
Procyanidin B2	PB2	29 (2) a	41 (5) ab	47 (3) b	40 (4) ab	54 (7) b
Phloretin	PHL	<0.05*	<0.05*	<0.05*	<0.05*	<0.05*
Phloretin-2'-*O*-glucoside	PHL-GLU	2.1 (0.3) ab	2.3 (0.3) ab	2.4 (0.2) a	1.8 (0.2) ab	1.49 (0.09) b
(+)-Catechin	CAT	40 (5) a	48 (4) a	49 (5) a	45 (3) a	44 (3) a
(-)-Epicatechin	EPI	0.86 (0.06) a	0.89 (0.08) a	0.89 (0.09) a	0.81 (0.05) a	0.79 (0.09) a
Total		109 (3) a	135 (7) b	130 (3) b	125 (3) b	141 (10) b
ESI (+)						
Peonidin-3-*O*-glucoside	PEO-GLU	0.111 (0.008) a	0.081 (0.006) bd	0.092 (0.006) ab	0.18 (0.02) c	0.070 (0.006) d
Cyanidin	CYA	10.7 (0.7) ac	16 (1) b	17 (1) b	12 (1) c	14 (2) abc
Cyanidin-3-*O*-galactoside	CYA-GAL	<0.01*	<0.01*	<0.01*	0.20 (0.02)	<0.01*
Cyanidin-3-*O*-glucoside	CYA-GLU	12 (1) a	8.8 (0.6) a	10.2 (0.8) a	10.0 (0.7) a	9.5 (0.7) a
Pelargonidin-3-*O*-glucoside	PEL-GLU	343 (21) a	218 (21) bc	319 (38) ac	293 (24) a	161 (17) b
Pelargonidin-3-*O*-rutinoside	PEL-RUT	40 (3) a	29 (3) bc	39 (4) ac	44 (5) a	20 (4) b
Total		406 (18) a	272 (24) b	385 (36) a	359 (22) a	205 (20) b

*Method detection limit. **Method quantification limit.

Figure 1 illustrates the score plots of PC1 vs PC2 (**Fig. 1A**) and PC1 vs PC3 (**Fig. 1B**), both accounting for a cumulative EV% >50%, as well as the corresponding loading plots (**Fig. 1C** and **Fig. 1D**). In both score plots QCs were very close to the origin of the coordinates, indicating the high accuracy and precision of the entire analytical procedure. Moreover, in both graphs, replicated samples showed quite similar score values, thus evidencing the homogeneous results obtained within each treatment. The five investigated samples were well discriminated in the PC1 vs PC2 space (EV%=59.1), thus highlighting the different influence exerted by irrigation waters on the expression of the phenolic secondary metabolism of strawberries (**Fig. 1A**). More in detail, the separation of TWW3 samples was due to their positive and high scores on PC2 and especially PC1, which are in turn related to CHL, QUE-GAL, and CYA-GAL concentrations (**Fig. 1C**). In this regard, it should be noted that these analytes were detected only in strawberries irrigated with TWW3. The higher PEO-GLU concentrations found in TWW3 samples also contributed to differentiate them from the others on the PC1 vs PC2 score plot. For TWW1 and TWW4 samples, which showed very similar coordinates on PC1, the separation was mainly due to their very different concentrations of KAM-RUT and PHL-GLU, the only two compounds providing very high loadings (in absolute value) on PC2 (**Fig. 1C**). TW and TWW2 were the closest samples in the PC1 vs PC2 score plot, reflecting the quite similar concentrations of most phenolic compounds in fruits from these treatments. The separation of TW and TWW2 samples was more evident when PC1 was plotted as a function of PC3 (EV%=54.0) (**Fig. 1B**). In fact, on this latter component, strawberries irrigated with TWW2 strongly differentiated from those grown with TW, due to concentration trends found in these samples for FER, PB2, CAT and CYA, all of them providing high absolute values of loadings in PC3 (**Fig. 1D**). Conversely, TWW1 and TWW4 samples grouped together, evidencing very similar behaviours of the two treatments on PC3 as well. According to the loading plot shown in **Fig. 1D**, fruits irrigated by TWW1 and TWW4 were characterized by high concentrations of QUE-GLU, PB1 and QUE-RUT, compared to those

found in the other samples. It is interesting to note that samples obtained with TWW1 and TWW4, which are effluents of WWTPs operating on similar mixed domestic-textile wastewaters and characterized by analogous treatments stages, exhibited very similar score values on both PC1 (EV%=38.7) and PC3 (EV%=15.3). The use of CA, as performed by using the squared Euclidean distances of autoscaled concentrations of targeted analytes (**Figure 2**), confirmed the homogeneous results obtained for replicated samples within each treatment. In fact, the replicates of each treatment grouped at very high similarity levels (i.e. TW=77.8%, TWW1=79.5%, TWW2=81.7%, TWW3=85.3%, and TWW4=81.1%), which were much greater than those regarding the other clusters present in the dendrogram (i.e. ≤51.3%).
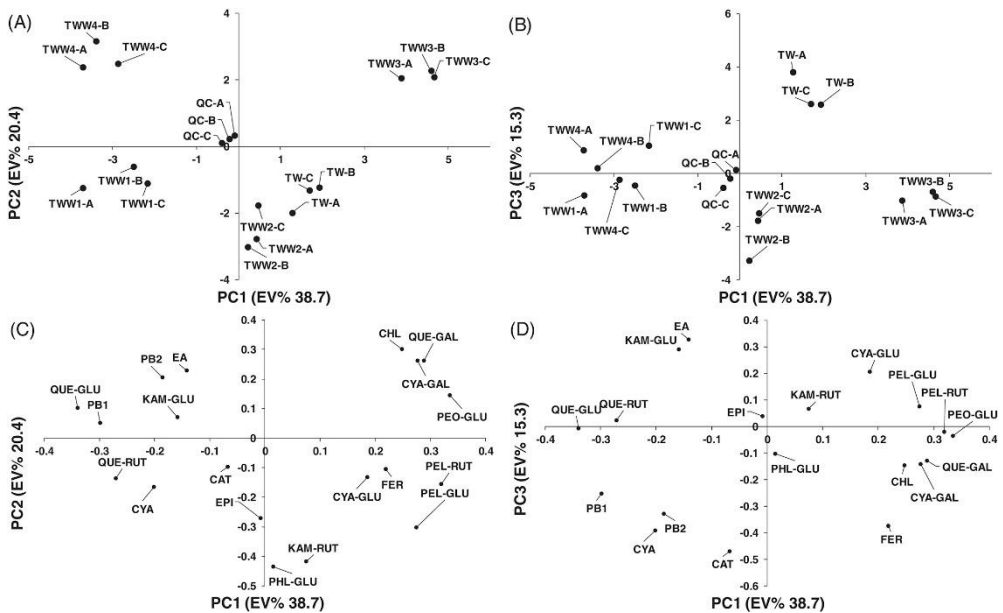
Figure 1 – Score (A,B) and loading (C,D) plots of PC1 versus PC2 (EV% = 59.1) and PC1 versus PC3 (EV% = 54.0). PCA values were calculated using autoscaled concentration values of target analytes. EV% = percentage of explained variance. The abbreviations used in the loading plots are defined in Table 3.
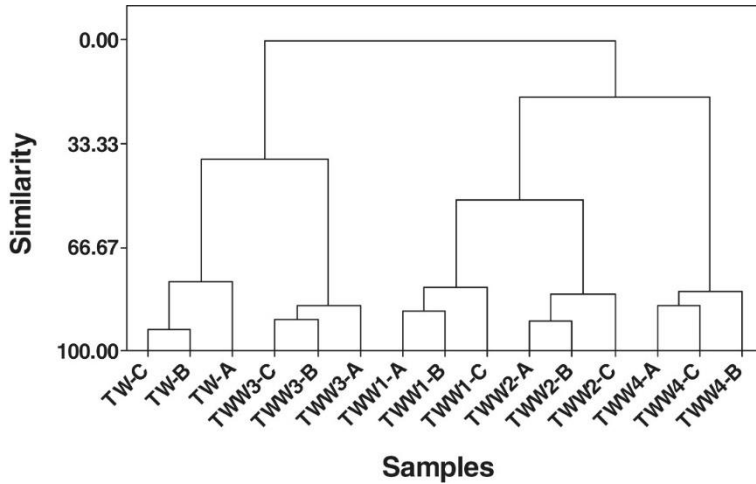
Figure 2 – Dendrogram of similarity of the 15 investigated strawberry samples, calculated on the basis of squared Euclidean distances of autoscaled concentration values of the 19 phenolic compounds detected in Camarosa fruits grown under irrigation with tap water (TW) and the four different treated wastewaters (TWW1, TWW2, TWW3 and TWW4). Uppercase letters A, B and C refer to the analysis of independent samples obtained with the same TWW.

2.4 Conclusions

Strawberry was a responsive fruit model for investigating the effect of irrigation with TWWs on fruit quality, which is an important aspect, currently not yet investigated in-depth, of the issue of wastewater reuse in agriculture. The comparative evaluation of the effect of various TWWs characterized by different physicochemical and chemical properties, allowed for obtaining interesting information that to the best of our knowledge are provided herein for the first time. Plants grown with TWW1 appeared to be among the most affected by non-conventional irrigation, displaying the lowest yield, sugar and TSP concentrations, RSA and AA values, as well as statistically lower TMA content, compared to control. Interestingly, strawberries irrigated with TWW2 and TWW3, which have common origin and underwent similar depuration stages, exhibited equivalent quality attributes. Fruits produced by irrigating plants with TWW4 showed erratic trends, being among the best for some parameters (e.g. yield) and among the worst for some others (e.g. TMA). This research also investigated for the first time a wide number of individual phenolic compounds as

quality indicators of non-conventional irrigation strategies, providing further important information. Concentrations of phenolic acids, flavonols and flavanols slightly increased with worsening the quality of TWW used for irrigation, whereas anthocyanins showed in almost all cases an opposite trend. Overall, these results showed that nutritional and nutraceutical attributes of strawberry fruits are strongly related to the quality of the water used for irrigation. However, the nutritional and nutraceutical attributes of the fruits obtained by non-conventional irrigation seem to be in line with strawberry marketability, even considering the fruits with the lowest quality attributes. It is remarkable that these results have been obtained by using TWWs with high SAR and conductivity values and chloride concentrations more than double than the maximum recommended for reuse in agriculture, and the Camarosa cultivar, which is considered very sensitive to the salinity of the irrigation water. In this regard, the presence in TWWs of significant concentrations of nutrients may has played an important role in the achievement of fruit nutritional and nutraceutical quality similar to the one elsewhere observed for strawberries grown under conventional irrigation. Accordingly, the reuse of TWWs in the agricultural sector may represent a valuable strategic solution for water saving (especially in countries experiencing water scarcity) suitable to increase the sustainability of soilless agricultural production, without losing fruit quality attributes and in full accordance with the principles of circular economy.

References

[1] V.C. Chaves, E. Calvete, F.H. Reginatto, Quality properties and antioxidant activity of seven strawberry (Fragaria x ananassa duch) cultivars, Scientia horticulturae, 225 (2017) 293-298 https://doi.org/10.1016/j.scienta.2017.07.013.
[2] M.F. Jaramillo, I. Restrepo, Wastewater reuse in agriculture: A review about its limitations and benefits, Sustainability, 9 (2017) 1734 https://doi.org/10.3390/su9101734.
[3] S. Bedbabis, B.B. Rouina, A. Mazzeo, G. Ferrara, Irrigation with treated wastewater affected the minor components of virgin olive oil from cv. Chemlali in Tunisia, European Food Research and Technology, 243 (2017) 1887-1894 https://doi.org/10.1007/s00217-017-2896-0.

[4] F. Pedrero, S. Camposeo, B. Pace, M. Cefola, G.A. Vivaldi, Use of reclaimed wastewater on fruit quality of nectarine in Southern Italy, Agricultural Water Management, 203 (2018) 186-192 https://doi.org/10.1016/j.agwat.2018.01.029.

[5] FAOSTAT, Food and Agriculture Organization of United Nations, 2017.

[6] G.D. Perulli, K. Bresilla, L. Manfrini, A. Boini, G. Sorrenti, L.C. Grappadelli, B. Morandi, Beneficial effect of secondary treated wastewater irrigation on nectarine tree physiology, Agricultural Water Management, 221 (2019) 120-130 https://doi.org/10.1016/j.agwat.2019.03.007.

[7] D. Bixio, C. Thoeye, J. De Koning, D. Joksimovic, D. Savic, T. Wintgens, T. Melin, Wastewater reuse in Europe, Desalination, 187 (2006) 89-101 https://doi.org/10.1016/j.desal.2005.04.070.

[8] Ministerial Decree 185/2003 of Italian Republic, "Regulation containing technical standards for wastewater reuse in implementation of article 26, paragraph 2, of Legislative Decree 152/1999", (2003).

[9] A. Christou, G. Maratheftis, E. Eliadou, C. Michael, E. Hapeshi, D. Fatta-Kassinos, Impact assessment of the reuse of two discrete treated wastewaters for the irrigation of tomato crop on the soil geochemical properties, fruit safety and crop productivity, Agriculture, ecosystems & environment, 192 (2014) 105-114 https://doi.org/10.1016/j.agee.2014.04.007.

[10] Z.H. Wang, S.X. Li, S. Malhi, Effects of fertilization and other agronomic measures on nutritional quality of crops, Journal of the Science of Food and Agriculture, 88 (2008) 7-23 https://doi.org/10.1002/jsfa.3084.

[11] G. Cirelli, S. Consoli, F. Licciardello, R. Aiello, F. Giuffrida, C. Leonardi, Treated municipal wastewater reuse in vegetable production, Agricultural Water Management, 104 (2012) 163-170 https://doi.org/10.1016/j.agwat.2011.12.011.

[12] S. Bedbabis, D. Trigui, C.B. Ahmed, M.L. Clodoveo, S. Camposeo, G.A. Vivaldi, B.B. Rouina, Long-terms effects of irrigation with treated municipal wastewater on soil, yield and olive oil quality, Agricultural Water Management, 160 (2015) 14-21 https://doi.org/10.1016/j.agwat.2015.06.023.

[13] A. Christou, G. Maratheftis, M. Elia, E. Hapeshi, C. Michael, D. Fatta-Kassinos, Effects of wastewater applied with discrete irrigation techniques on strawberry plants' productivity and the safety, quality characteristics and antioxidant capacity of fruits, Agricultural Water Management, 173 (2016) 48-54 https://doi.org/10.1016/j.agwat.2016.04.027.

[14] F. Giampieri, S. Tulipani, J.M. Alvarez-Suarez, J.L. Quiles, B. Mezzetti, M. Battino, The strawberry: composition, nutritional quality, and impact on human health, Nutrition, 28 (2012) 9-19 https://doi.org/10.1016/j.nut.2011.08.009.

[15] C.K. Chandler, K. Folta, A. Dale, V.M. Whitaker, M. Herrington, Strawberry, Fruit breeding, Springer 2012, pp. 305-325.

[16] D.L. Suarez, C.M. Grieve, Growth, yield, and ion relations of strawberry in response to irrigation with chloride-dominated waters, Journal of plant nutrition, 36 (2013) 1963-1981 https://doi.org/10.1080/01904167.2013.766210.

[17] M. Mikulic-Petkovsek, V. Schmitzer, A. Slatnar, F. Stampar, R. Veberic, Composition of sugars, organic acids, and total phenolics in 25 wild or cultivated

berry species, Journal of food science, 77 (2012) C1064-C1070 https://doi.org/10.1111/j.1750-3841.2012.02896.x.

[18] J. Milivojević, V. Maksimović, M. Nikolić, J. Bogdanović, R. Maletić, D. Milatović, Chemical and antioxidant properties of cultivated and wild Fragaria and Rubus berries, Journal of Food Quality, 34 (2011) 1-9 https://doi.org/10.1111/j.1745-4557.2010.00360.x.

[19] A.B. Cerezo, E. Cuevas, P. Winterhalter, M. Garcia-Parrilla, A. Troncoso, Isolation, identification, and antioxidant activity of anthocyanin compounds in Camarosa strawberry, Food Chemistry, 123 (2010) 574-582 https://doi.org/10.1016/j.foodchem.2010.04.073.

[20] M. Del Bubba, L. Checchini, U. Chiuminatto, S. Doumett, D. Fibbi, E. Giordani, Liquid chromatographic/electrospray ionization tandem mass spectrometric study of polyphenolic composition of four cultivars of Fragaria vesca L. berries and their comparative evaluation, Journal of Mass Spectrometry, 47 (2012) 1207-1220 https://doi.org/10.1002/jms.3030.

[21] G. La Barbera, A.L. Capriotti, C. Cavaliere, S. Piovesana, R. Samperi, R.Z. Chiozzi, A. Laganà, Comprehensive polyphenol profiling of a strawberry extract (Fragaria× ananassa) by ultra-high-performance liquid chromatography coupled with high-resolution mass spectrometry, Analytical and bioanalytical chemistry, 409 (2017) 2127-2142 https://doi.org/10.1007/s00216-016-0159-8.

[22] Publiacqua S.p.A., Monitoraggio sistemi acquedottistici (in Italian), 2019.

[23] S. Doumett, D. Fibbi, A. Cincinelli, E. Giordani, S. Nin, M. Del Bubba, Comparison of nutritional and nutraceutical properties in cultivated fruits of Fragaria vesca L. produced in Italy, Food Research International, 44 (2011) 1209-1216 https://doi.org/10.1016/j.foodres.2010.10.044.

[24] C.V.A. Scordo, L. Checchini, L. Renai, S. Orlandini, M.C. Bruzzoniti, D. Fibbi, L. Mandi, N. Ouazzani, M. Del Bubba, Optimization and validation of a method based on QuEChERS extraction and liquid chromatographic–tandem mass spectrometric analysis for the determination of perfluoroalkyl acids in strawberry and olive fruits, as model crops with different matrix characteristics, Journal of Chromatography A, 1621 (2020) 461038 https://doi.org/10.1016/j.chroma.2020.461038.

[25] C. Ancillotti, L. Ciofi, D. Pucci, E. Sagona, E. Giordani, S. Biricolti, M. Gori, W.A. Petrucci, F. Giardi, R. Bartoletti, Polyphenolic profiles and antioxidant and antiradical activity of Italian berries from Vaccinium myrtillus L. and Vaccinium uliginosum L. subsp. gaultherioides (Bigelow) SB Young, Food chemistry, 204 (2016) 176-184 https://doi.org/10.1016/j.foodchem.2016.02.106.

[26] W. Brand-Williams, M.-E. Cuvelier, C. Berset, Use of a free radical method to evaluate antioxidant activity, LWT-Food science and Technology, 28 (1995) 25-30 https://doi.org/10.1016/S0023-6438(95)80008-5.

[27] R. Re, N. Pellegrini, A. Proteggente, A. Pannala, M. Yang, C. Rice-Evans, Antioxidant activity applying an improved ABTS radical cation decolorization assay, Free radical biology and medicine, 26 (1999) 1231-1237 https://doi.org/10.1016/S0891-5849(98)00315-3.

Section 2

[28] I.F. Benzie, J.J. Strain, The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay, Analytical biochemistry, 239 (1996) 70-76 https://doi.org/10.1006/abio.1996.0292.

[29] T.A. Bauder, R. Waskom, P. Sutherland, J. Davis, R. Follett, P. Soltanpour, Irrigation water quality criteria, Service in action; no. 0.506, (2011).

[30] J.F. Ferreira, X. Liu, D.L. Suarez, Fruit yield and survival of five commercial strawberry cultivars under field cultivation and salinity stress, Scientia Horticulturae, 243 (2019) 401-410 https://doi.org/10.1016/j.scienta.2018.07.016.

[31] I. Maksimović, M. Putnik-Delić, I. Gani, J. Marić, Ž. Ilin, Growth, ion composition, and stomatal conductance of peas exposed to salinity, Central European Journal of Biology, 5 (2010) 682-691 https://doi.org/10.2478/s11535-010-0052-y.

[32] P. Parihar, S. Singh, R. Singh, V.P. Singh, S.M. Prasad, Effect of salinity stress on plants and its tolerance strategies: a review, Environmental Science and Pollution Research, 22 (2015) 4056-4075 https://doi.org/10.1007/s11356-014-3739-1.

[33] C. An, H. Yoon, Y. Hwang, H.J. Hwang, G.M. Shon, Effects of reuse of organic substrate on growth and yield of strawberry in soilless culture, XXVII International Horticultural Congress-IHC2006: International Symposium on Advances in Environmental Control, Automation 761, 2006, pp. 521-526 https://doi.org/10.17660/ActaHortic.2007.761.72.

[34] C. Diara, L. Incrocci, A. Pardossi, A. Minuto, Reusing greenhouse growing media, Acta Horticulturae, 927 (2012) 793-800 https://doi.org/10.17660/ActaHortic.2012.927.98.

[35] I. Akhatou, Á. Fernández Recamales, Influence of cultivar and culture system on nutritional and organoleptic quality of strawberry, Journal of the Science of Food and Agriculture, 94 (2014) 866-875 https://doi.org/10.1002/jsfa.6313.

[36] F. Tozzi, M. Del Bubba, W.A. Petrucci, S. Pecchioli, C. Macci, F.H. García, J.J.M. Nicolás, E. Giordani, Use of a remediated dredged marine sediment as a substrate for food crop cultivation: Sediment characterization and assessment of fruit safety and quality using strawberry (Fragaria x ananassa Duch.) as model species of contamination transfer, Chemosphere, 238 (2020) 124651 https://doi.org/10.1016/j.chemosphere.2019.124651.

[37] I. Akhatou, A.n. Fernández-Recamales, Nutritional and nutraceutical quality of strawberries in relation to harvest time and crop conditions, Journal of agricultural and food chemistry, 62 (2014) 5749-5760 https://doi.org/10.1021/jf500769x.

[38] E. Kafkas, M. Koşar, S. Paydaş, S. Kafkas, K. Başer, Quality characteristics of strawberry genotypes at different maturation stages, Food Chemistry, 100 (2007) 1229-1236 https://doi.org/10.1016/j.foodchem.2005.12.005.

[39] A. Keutgen, E. Pawelzik, Modifications of taste-relevant compounds in strawberry fruit under NaCl salinity, Food chemistry, 105 (2007) 1487-1494 https://doi.org/10.1016/j.foodchem.2007.05.033.

[40] Y. Awang, J. Atherton, A. Taylor, Salinity effects on strawberry plants grown in rockwool. II. Fruit quality, Journal of Horticultural Science, 68 (1993) 791-795 https://doi.org/10.1080/00221589.1993.11516414.

[41] C. Ubeda, R. Callejón, C. Hidalgo, M. Torija, A. Troncoso, M. Morales, Employment of different processes for the production of strawberry vinegars: Effects on antioxidant activity, total phenols and monomeric anthocyanins, LWT-Food Science and Technology, 52 (2013) 139-145 https://doi.org/10.1016/j.lwt.2012.04.021.

[42] M. Alonzo-Macías, A. Cardador-Martínez, S. Mounir, G. Montejano-Gaitán, K. Allaf, Comparative study of the effects of drying methods on antioxidant activity of dried strawberry (Fragaria Var. Camarosa), Journal of Food Research, 2 (2013) 92-107 http://dx.doi.org/10.5539/jfr.v2n2p92.

[43] D. Hernanz, Á.F. Recamales, A.J. Meléndez-Martínez, M.L. González-Miret, F.J. Heredia, Assessment of the differences in the phenolic composition of five strawberry cultivars (Fragaria× ananassa Duch.) grown in two different soilless systems, Journal of agricultural and food chemistry, 55 (2007) 1846-1852 https://doi.org/10.1021/jf063189s.

3 **Liquid chromatographic quadrupole time-of-flight mass spectrometric untargeted profiling of (poly)phenolic compounds in *Rubus idaeus* L. and *Rubus Occidentalis* L. fruits and their comparative evaluation.**

Section 3

Abstract

This study provided a detailed profiling of the antioxidant and bioactive compounds occurring in three varieties of *Rubus idaeus* L. fruits ("Fall Gold", "Glen Ample" and "Tulameen") compared to *Rubus occidentalis* L. black raspberry ("Jewel" cultivar), adopting a comprehensive untargeted metabolomics approach developed with UHPLC analysis coupled with quadrupole/time-of-flight high resolution mass spectrometry, using the SWATH® acquisition protocol. The feature selection and annotation workflow, applied to the analysis of raspberry extracts in both polarities, allowed identifying 68 bioactive compounds mainly belonging to the classes of (poly)phenolic compounds. Interestingly, some of these identifications (e.g., ferulic acid glycosides and the ellagitannin-like nobotanin/malabathrin) represent the first report in raspberry fruits. Principal component analysis made possible highlighting the features more related to the expression of a genotype effect within the *R. idaeus* species or between the two raspberry species herein investigated. Overall, flavanols were the most discriminating features for the Fall Gold variety, whereas ellagitannins and flavonol glycosides represent more distinctive metabolic traits in Glen Ample and Tulameen fruits. Moreover, *R. occidentalis* Jewel variety was strongly characterized by the occurrence of anthocyanins, such as cyanidin, pelargonidin and delphinidin glycosides.

Keywords: raspberry; metabolomics; high resolution mass spectrometry; data independent acquisition; phenolic compounds; pomological characterization

3.1 Introduction

Wild berries are well-known sources of bioactive compounds, mainly phenolics, synthesized by the plant secondary metabolism [1-3]. Such native chemicals play a crucial role in the prevention of a wide range of chronic and degenerative diseases [4,5]. For example, bilberry, blueberry, and raspberry extracts demonstrated to exert antiproliferative and/or pro-apoptotic effects on specific cancer cell lines [6-8], whereas the intake of blueberry and strawberry

supplements showed beneficial effects in the prevention and treatment of cardiovascular diseases [9] and diabetes [10], respectively. Within the wide variety of wild berries, raspberry and blackberry are the most consumed edible fruits of a multitude of plant species in the genus *Rubus* of the *Rosaceae* family [11]. *Rubus* fruits are considered important functional foods due to their remarkable nutritional value and content of dietary antioxidants [11,12]. Within *Rubus* species, blackberry is commonly harvested as spontaneous wild berry, whilst raspberry is widely cultivated in Europe and North America as *Rubus idaeus* L. and *Rubus occidentalis* L. species, respectively [13], with a worldwide annual production of about one million tons in 2019 [14]. In spite of their significant commercial value and recognized dietary importance, raspberry fruits have been investigated for their bioactive phenolic constituents with a quite limited extent. More in detail, the phenolic composition of *R. occidentalis* berries has been poorly described in the literature, since to our knowledge only three studies have been published so far [2,15,16], whilst a few more studies have been performed on *R. idaeus* fruits [2,17-23]. Some of these studies have been performed using liquid chromatography (LC) hyphenated with diode array detection and single quadrupole mass spectrometry through a selected ion monitoring approach, highlighting the occurrence of anthocyanins (i.e. cyanidin and pelargonidin derivatives) [2,19] and 32 compounds belonging to phenolic acids, flavanols, flavonols, ellagic acid and its derivatives, and ellagitannins [2]. Other researches [17,18,20,21] adopted LC coupled with linear ion trap low-resolution mass spectrometer, which is in principle capable of more-in-depth structural analyses compared to single quadrupole. In this regard, a significant improvement of the identification process was obtained by Dincheva and co-workers [18], thus pushing the annotation up to 60 compounds belonging to the flavonols, flavanols, and hydroxycinnamic acids, besides anthocyanins and ellagitannins. However, the use of low-resolution mass spectrometry does not provide the exact mass of both precursor and fragment ions, thus resulting in a low-accuracy identification, unless authentic standards are available. Conversely,

high-resolution mass spectrometry, especially in the tandem mode (e.g. quadrupole/time-of-flight or quadrupole-orbital trap), is much more informative, since it allows for obtaining important structurally related information through the fragmentation of parent molecules and the accurate mass readout of both precursors and fragments [1,24]. This consideration highlights the importance of adopting acquisition strategies based on high-resolution mass spectrometry, which has been applied in few cases for the exploration of the phenolic fraction of raspberry fruits, using LC hyphenated with time-of-flight [22] or quadrupole/time-of-flight (Q/TOF) [15,23] mass spectrometry (MS). More in detail, untargeted LC-Q/TOF was recently adopted for investigating negatively ionisable phenolic compounds in *R. occidentalis* berries [15]. Despite the accurate workflow adopted for feature extraction and annotation, the list of identified analytes was limited to a small number of compounds belonging to phenolic acids, flavonols, flavanols, ellagic acid and ellagic acid glycosides. A deeper untargeted analysis was performed by Zhang and co-workers [23] on *R. idaeus* extracts, identifying 50 phenolic secondary metabolites included in the group of anthocyanins and in the other aforementioned categories. The expression of phenolic compounds in *R. idaeus* raspberries has been investigated as a function of stage of maturity [17], growing region [22], and genotype [17,22,25], suggesting the last one as the variable mostly influencing the phenolic profile. However, no data are reported in literature concerning the effect of genotype between the R. idaeus and R. occidentalis berry species.

Based on the aforementioned considerations, this paper aims at (i) deepening the knowledge of the phenolic fraction of R. idaeus and R. occidentalis berries, the latter being poorly described in literature, and (ii) investigating the genotype effect on the phenolic expression in the two species. To these purposes, an untargeted metabolomics LC-Q/TOF study was performed on berries from "Glen Ample", "Tulameen", and "Fall Gold" varieties of R. idaeus, and from "Jewel" cultivar of R. occidentalis, cultivated in the same orchard and under the same rural practices.

3.2 Materials and Methods

3.2.1 Reagents and Standards

Analytical reference standards for identity confirmation were supplied, as specified in Section 1 of the *Supplementary Materials*. Hydrochloric acid (37%), LC-MS grade methanol and water were obtained from J.T. Baker (Deventer, the Netherlands). HPLC grade acetone, glacial acetic acid and LC-MS formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). Sodium fluoride, Folin–Ciocalteu reagent and sodium carbonate were obtained by Merck (Darmstadt, Germany). Ultrapure water was taken from a Milli-Q system (Millipore, Billerica, MA, USA). Nylon membranes (porosity 0.2 μm) for the filtration of raspberry extracts before HPLC analysis were obtained from VWR™ International (Radnor, PA, USA).

3.2.2 Sample Preparation

R. idaeus and *R. occidentalis* analyzed in the present study were grown in a same experimental site located in Tuscany (44°02.175′, 10°47.361′, altitude 450 m a.s.l.), under the same agricultural conditions. The cultivars of R. idaeus investigated were "Glen Ample" (GA, red), "Tulameen" (T, red) and "Fall Gold" (G, yellow), whereas, for R. occidentalis, the cultivar "Jewel" (J, black) was selected. Six independent samples, each consisting of ten fresh fruits free of defects, were prepared for each raspberry cultivar. Three samples were used for the untargeted LC-Q/TOF study, while the others were destined to the pomological analyses. After the sampling, the berries intended for extraction were immediately frozen in liquid nitrogen, freeze-dried, grinded to obtain a homogeneous powder and finally stored at −20 °C until extraction and analysis were performed. Three representative aliquots from each freeze-dried berry sample were extracted according to a procedure previously developed by Zhang and co-workers [23] with few modifications. Briefly, about 500 mg dry weight (d.w.) raspberry aliquots were mixed with 5 mL of acetone/10 mM NaF water/acetic acid (70:29.7:0.3, v/v/v) and then vortexed for 30 s, sonicated in ice

bath at controlled temperature (~0 °C) for 5 min and centrifuged at 8000× g for 10 min. This extraction protocol was replicated four times and its efficiency evaluated by measuring spectrophotometrically total soluble polyphenols (TSP) and total monomeric anthocyanins (TMA), as reported by Renai et al. [26] (see Section 2 of the *Supplementary Materials* for full details), using calibration curve ranges of 5–10 µg of procyanidin B1 and 20–300 µg of cyanidin-3-O-sophoroside, respectively. The results show that the fourth extraction allowed for recovering 3.6–6.5% and 3.0–5.8% of TSP and TMA determined with three extractions. Accordingly, for untargeted LC-Q/TOF analysis, three sequential extractions were performed on three independent aliquots of each raspberry cultivar and the resulting extracts were combined. The organic solvent was removed by vacuum evaporation and filtered at 0.2 µm using nylon membranes, and the resulting aqueous extract was analyzed.

3.2.3   *Pomological analyses*

Colorimetric coordinates (i.e., L, a, b), total soluble solids (TSS) and titratable acids (TA) were performed according to procedures elsewhere reported [27,28], in order to evaluate the stage of maturity of the harvested fruits. In detail, colorimetric coordinates were determined with a Minolta Chromameter CR200 (Konica Minolta, Tokyo, Japan) electronic colorimeter equipped with a pulsed xenon arc lamp inside a mixing chamber, which provides diffuse, uniform lighting over the 8-mm-diameter specimen area. TSS were determined using an Atago N1 digital refractometer (Atago Co., Ltd., Tokyo, Japan) and expressed as a percentage (°Brix). TA were determined by automatic titration (877 Titrino plus, Metrohm) with 0.1 M solution of NaOH up to pH 8.1, and the results were expressed as g of citric acid (CA) per kg of fresh fruit (f.w.). The results obtained are reported in **Table 1**.

3.2.4   *LC-Q/TOF Analysis*

LC analysis was performed on an ExionLC analytical UHPLC system (SCIEX, Framingham, MA, USA) equipped with an Acquity BEH C18 column (15 cm ×

2.1 cm i.d., particle size 1.7 μm) and a guard column containing the same stationary phase (Waters, Milford, MA, USA). Column temperature was set at 50 °C. LC-MS water (eluent A) and methanol (eluent B) solutions were used for the analyte elution, acidifying each solvent with formic acid at 0.1% and 5% (v/v) for negative and positive electrospray ionization (ESI), respectively. The following elution gradient was adopted: 0–3 min, isocratic 2% B; 3–35 min, linear gradient 2–100% B; and 35–38 min, isocratic 100% B. The flow rate was 450 μL/min and the injection volume was 5 μL. The LC system was coupled with a TripleTOF® 6600 Q/TOF mass analyzer (SCIEX, Framingham, MA, USA) by the DuoSpray™ Source for TOF and Q/TOF analyses. The following source parameters were used during the acquisitions; (i) positive ionization: heater temperature 450 °C, Curtain Gas™ 30, nebulizing gas 55, heating gas 65 and spray voltage +5000 V; and (ii) negative ionization: heater temperature 450 °C, Curtain Gas™ 30, nebulizing gas 45, heating gas 55 and spray voltage −4500 V. Under both ionization modes, each extract was analyzed using the SWATH™ data independent acquisition protocol, which allows simultaneously acquiring TOF full scan MS and Q/TOF MS2 spectra with a comprehensive detection approach, i.e., virtually detecting all the analytes present in the extract and eluting under the chromatographic conditions adopted. The high-resolution TOF MS full scan experiment was carried out in the range 100–1000 Da (cycle time 200 ms), with an accumulation time of 150 ms and a collision energy of 70 eV. Automated calibration was performed by an external calibrant delivery system (CDS), infusing proper calibration solution prior to sample introduction.

3.2.5   *Data Processing for Feature Selection and Identification*

The number of raw data derived from the SWATH™ analysis of the investigated samples is very high, thus needing to be processed with specific software. In this study, Marker View® 1.3.1 software was used for instrumental noise removal and blank subtraction, spectra deconvolution and chromatogram alignment, based on the TOF exact mass and isotope pattern determinations, as well as on the Q/TOF fragmentation spectra of parent ions. The following alignment criteria were

adopted within the three independent samples of each cultivar: (i) TOF accuracy of the pseudo-molecular ion < 5 ppm; (ii) isotope ratio difference compared to the theoretical isotope profile < 20%; (iii) purity score of the MS2 spectra compared to the one of available standards ≥ 80%; and (iv) retention time (tR) tolerance ≤ 0.05 min [1]. Afterwards, the following workflow was adopted using the R software (version 4.0.3, https://cran.r-project.org, accessed on 1 March 2021) for the selection of a restricted group of the aligned features. In detail, the following were selected: (i) features with a coefficient of variation (CV) among cultivars higher than that in QCs; (ii) features that after the Kruskal–Wallis comparison among cultivars were significantly different at the probability level of 2% (p-value < 0.02); and (iii) features that after the post-hoc Dwass, Steel, Critchlow-Fligner (DSCF) multiple comparison analysis, based on pairwise two-sample Wilcoxon comparisons [29] among sample groups were significantly different at the probability level of 20% (p-value < 0.2). Manual revision of the features resulting from the aforementioned workflow was finally carried out with the purpose of retaining only the features providing acceptable spectral data for the successive annotation. In this study, according to metabolomics guidelines, four levels of feature annotation were distinguished [30]. Briefly, Level I was assigned when the feature in the aligned chromatogram was successfully compared with the reference standard. For Level II annotation, the identification is performed based upon exact mass value, isotopic profile, MS2 spectra and chromatographic behavior of the aligned feature in comparison with internal and/or freely available external libraries (i.e., MassBank, GNPS, MetaboBASE, Fiehn/Vaniya natural product library and BMDMS-NP) of mass spectra and literature information (putatively annotated compound). The identification at Level III is based on characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class (putatively characterized compound classes). Finally, Level IV includes the unknown compounds.

3.2.6 *Chemometrics Analyses for Genotype Comparison*

To compare the metabolome profiles of the four investigated cultivars and to highlight the features that mainly contributed to their differentiation, principal component analysis (PCA) of molecular or quasi-molecular ions of the assigned compounds was performed using MarkerView software. This approach was carried out separately for compounds detected under negative and positive ionization modes. Quality control (QC) of PCA was performed, using QC samples (n = 3), which consisted of a mixture of equal aliquots of each raspberry extract. QC evaluation was carried out by verifying if PCA object scores obtained by replicated injections of the QC sample were close to the origin of PCA coordinates. Pearson product-moment correlation (PPMC) analysis was also performed to evaluate the grade of correlation among the selected features and pomological data.

3.3   Results and Discussion

3.3.1   *Pomological Parameters*

Pomological parameters are illustrated in **Table 1**. The CIELAB spatial coordinates measured on raspberry skin were in the ranges of 25–53, 2–23 and 3–34 for brightness (L), green to red (a) and blue to yellow (b) axes, respectively. These very wide variations obviously reflected the different skin colors of the investigated berries. TSS and TA showed not negligible variations, being in the ranges 8.1–13.1° Brix and 10.7–19.9 g of citric acid per kg of fresh weight fruit. However, the two red raspberry cultivars (R. idaeus GA and T) exhibited the same CIELAB coordinates; moreover, for these two genotypes, statistically comparable values of TSS and TA were found, thus highlighting for these fruits the same stage of maturity. It should also be noted that the values of pomological parameters found here for GA, T, G and J were in quite good agreement with data previously reported for fully mature raspberries of the same cultivars [31,32,33,34] and/or for other genotypes [35], thus indicating that the harvest time was properly chosen.

Table 1 – Fruit pomological characterization of the investigated raspberry samples. Values of colorimetric coordinates (L, a and b), total soluble solids (TSS, expressed as Brix) and titratable acids (TA) are reported as means (n = 3) and relative standard deviations (in brackets). GA, Glen Ample; G, Fall Gold; T, Tulameen; J, Jewel; CA, citric acid; f.w., fruit fresh weight. Values marked with a different superscript letter (a, b, c) are statistically different (p-value < 0.05) according to Kruskal–Wallis Test.

	L	a	b	TSS(°Brix)	TA (g CA kg^{-1} f.w.)
Rubus idaeus					
GA	35 (1)[a]	23 (3)[a]	16 (3)[a]	11.1 (1.1)[a]	19.7 (1.4)[a]
T	35 (3)[a]	23 (3)[a]	16 (3)[a]	11.8 (1.7)[a]	19.9 (1.3)[a]
G	53 (1)[b]	12 (2)[a]	34 (2)[b]	8.1 (1.1)[b]	12.4 (1.0)[b]
Rubus occidentalis					
J	25 (1)[c]	2 (1)[b]	3 (1)[c]	13.1 (1.1)[c]	10.7 (1.0)[c]

3.3.2 *Feature Selection and Annotation*

Data peak picking and retention time alignment resulted in a very large number of features in both negative and positive ionization modes, i.e., 13,211 and 44,251, respectively. The further workflow of feature selection provided an additional restriction of the aligned data, which accounted for 10,323 and 43,774 for ESI(−) and ESI(+), respectively. Non-parametric Kruskal–Wallis test highlighted 1732 features in negative polarity and 3088 features in positive polarity. Finally, contrast analysis and manual revision of the significant features resulted in 49 and 19 features in negative and positive ionization, respectively. All selected features were annotated according to the identification criteria reported in Section 2.5. For each feature, **Table 2** and **Table 3** illustrate t_R (min), mass to charge ratio (*m/z*) of the pseudo-molecular or the molecular ion found by the full scan experiment, main mass fragments, proposed formula and the corresponding exact mass, the mass accuracy (Δ, ppm), the raspberry variety in which the metabolite was identified and the tentative identification.

Feature Annotation

Phenolic acids: Several phenolic acids belonging to the class of the hydroxybenzoic (Peaks 1 and 2) and hydroxycinnamic acids (Peaks 6, 9, 17, 18 and 27) were identified under negative ionization in at least one raspberry sample

(**Table 2**). In detail, Peaks 1 and 2, which were identified at Level I as gallic acid and 3,4-dihydroxybenzoic acid, respectively, were detected only in the *R. occidentalis* cultivar. Among hydroxycinnamic acids, chlorogenic acid (Peak 17) was annotated at Level I, whereas the other acids (i.e., ferulic, p-coumaric and sinapic), occurring only as hexose glycosides, were identified at Level II with $|\Delta|$ of 2.4–4.9 ppm. Interestingly, the two ferulic acid hexosides identified here in *R. idaeus* fruits (Peaks 6 and 9) represent the first report of ferulic acid glycosides in raspberry. As illustrated in **Figure S1** and **Scheme S1**, these peaks were identified at Level II through their $[M-H]^-$ pseudo-molecular ion at *m/z* 355.10, which fragmented, giving rise to peaks at *m/z* 191.02 (164 Da, loss of hexose), 147.03 and 129.02 (successive losses of CO2 and water, respectively), as already reported in other berries [1]. Peak 18 showed a $[M-H]^-$ pseudo-molecular ion at *m/z* 325.09, which fragmented in *m/z* 118.04 and 117.03 due to the cleavage of the anomeric bond (loss of 163 Da) and successive loss of CO_2 (loss of 44 Da). Based on these findings and previously reported patterns of fragmentation and related annotations [18,23], this feature was putatively identified (i.e., Level II) as a p-coumaroyl hexoside. Peak 27 with $[M-H]^-$ at *m/z* 385.12 was putatively annotated as sinapic acid hexoside since it fragmented giving rise to m/z 223.06 (loss of the hexose group), which further fragmented in *m/z* 205.05 (formal loss of H_2O, 18 Da) and 190.03 (loss of methyl group) (**Figure S2** and **Scheme S2**).

Ellagitannins: Thirteen ellagitannins (Peaks 3, 7, 8, 12, 19, 21–24 and 29) were identified under negative ionization in both *R. idaeus* and *R. occidentalis* raspberry (**Table 2**).

Table 2 – Retention times (t_R, min), pseudo-molecular ions (TOF, Da), mass fragments (Q/TOF, Da), proposed formula, corresponding exact mass (Da) and accuracy (Δ, ppm) of peaks tentatively identified under negative ionization. Symbols + and − mean detected and not detected, respectively. GA, Glen Ample; T, Tulameen; J, Jewel; G, Fall Gold. a Peaks annotated at Level I. The base peak of each MS2 spectra is reported in bold character. Fragments marked with asterisk are double-charged ions.

Peak	t_R	TOF	Charge	Q/TOF	Formula	Exact mass	Δ	GA	T	G	J	Tentative identification
1	1.61	169.0144	-H	**125.0233**; 126.0275; 124.0154; 79.0173	C7H6O5	169.0143	0.6	-	-	-	+	Gallic acid [a]
2	3.17	153.0200	-H	**109.0291**; 108.0210	C7H6O4	153.0194	3.9	-	-	-	+	3,4-Dihydroxybenzoic acid [a]
3	5.28	633.0735	-H	331.0685; **301.0003**	C27H22O18	633.0733	0.3	+	+	+	+	Galloyl-HHDP-hexose
4	6.03	577.1380	-H	425.0909; **407.0804**; 289.0732; 125.0232	C30H26O12	577.1352	4.9	+	+	+	+	B-type procyanidin dimer
5	6.26	865.2017	-H	**577.1394**; 575.1259; 287.0570	C45H38O18	865.1985	3.7	+	+	+	+	Procyanidin C1 [a]
6	6.34	355.1024	-H	191.0203; 147.0304; **129.0192**	C16H20O9	355.1035	-3.1	+	+	+	-	Ferulic acid hexoside I
7	6.49	783.0703	-2H	935.0892; 933.0731; 633.0777; 617.0370; 331.0679; **300.9993**	C68H48O44	783.0687	2.0	+	+	+	+	Sanguiin H10 I
8	6.68	858.0684	-2H	935.0890; 858.0752; 633.0756; 631.0607; **300.9991**	C75H50O48	858.0663	2.4	+	+	+	+	Sanguiin H6 degalloylated
9	6.89	355.1026	-H	191.0203; 147.0304; **129.0192**	C16H20O9	355.1035	-2.5	+	+	+	-	Ferulic acid hexoside II
10	7.06	289.0724	-H	245.0819; 205.0502; 203.0707; **125.0228**; 123.0434; 109.0276	C15H14O6	289.0718	2.2	+	+	+	+	(+)-Catechin [a]
11	7.13	863.1903	-H	711.1413; 693.1335; 575.1237; 449.0896; 423.0721; 413.0851; 405.0695; **287.0567**	C45H36O18	863.1914	-1.3	+	+	+	-	A/B type procyanidin trimer
12	7.31	633.0760	-H	**300.9993**	C27H22O18	633.0733	4.3	+	+	+	+	Corilagin
13	7.31	609.1490	-H	**300.0287**; 301.0351; 178.9983; 151.0035	C27H30O16	609.1461	4.7	-	-	-	+	Quercetin deoxyhexose-hexoside
14	7.46	577.1378	-H	407.0822; **289.0757**; 125.0233	C30H26O12	577.1352	4.6	+	+	+	+	Procyanidin B1
15	7.67	463.0892	-H	327.0522; **175.0255**; 125.0234	C21H20O12	463.0882	2.2	+	+	+	-	Tetrahydroxyflavonol-3-O-hexoside
16	7.85	865.2023	-H	**577.1380**; 407.0781; 287.0560; 125.0224	C45H38O18	865.1985	4.4	+	+	+	+	B-type procyanidin trimer
17	8.13	353.0884	-H	**191.0551**	C16H18O9	353.0878	1.7	+	+	+	+	Chlorogenic acid [a]
18	8.37	325.0940	-H	**146.0319**; 145.0289; 118.0364; 117.0332	C15H18O8	325.0929	3.4	+	+	+	+	p-Coumaryl hexoside
19	8.44	783.0702	-2H	935.0892; 933.0731; 633.0777; 617.0370; 331.0679; **300.9993**	C68H48O44	783.0687	1.9	+	+	+	+	Sanguiin H10 II
20	8.60	577.1355	-H	425.0887; 407.0775; **289.0716**	C30H26O12	577.1351	0.6	+	+	+	+	Procyanidin B2 [a]
21	9.20	933.7395	-3H	617.0367*; **300.9912**	C123H80O78	933.7358	4.0	+	+	+	+	Lambertianin C
22	9.25	934.0796	-2H	915.0632; 897.0499; 633.0775; **301.0077**	C82H54O52	934.0757	4.2	+	+	+	+	Sanguiin H6 I
23	9.57	934.0779	-2H	915.0618; 897.0485; 633.0759; **301.0056**	C82H54O52	934.0737	4.5	+	+	+	+	Sanguiin H6 II
24	9.58	551.0433	-2H	469.0072; **300.9998**; 169.0133	C48H32O31	551.0410	4.2	+	+	+	+	Sanguiin H2
25	9.82	289.0724	-H	245.0818; 203.0704; 125.0226; **123.0434**	C15H14O6	289.0717	2.2	+	+	+	+	(-)-Epicatechin [a]
26	10.23	341.1245	-H	**179.0710**; 121.0280	C16H22O8	341.1241	1.0	-	-	+	-	Coniferin
27	10.28	385.1154	-H	223.0622; **205.0528**; 190.0276	C17H22O10	385.1135	4.9	+	+	+	-	Sinapic acid hexoside
28	10.33	859.0802	-2H	785.0884; 633.0781; **300.9993**	C75H52O48	859.0760	4.9	+	+	+	-	Nobotanin/Malabathrin ET-like
29	10.61	593.1510	-H	475.1400; 431.0600; **245.1390**	C27H30O15	593.1506	-0.01	+	+	-	+	Apigenin diglucoside
30	11.69	651.1983	-H	**593.1694**; 325.0782; 285.0420; 284.0347	C30H38O17	651.1945	4.6	+	+	-	+	Trihydroxy-methoxyflavone deoxyhexose-hexose derivative

31	11.71	303.0523	-H	285.0417; **125.0233**	C15H12O7	303.0510	4.2	-	-	+	-	Taxifolin [a]
32	11.91	625.1433	-H	**301.0323**; 245.0937	C27H30O17	625.1410	3.7	+	+	+	-	Quercetin-3-O-sophoroside [a]
33	12.14	389.1244	-H	**227.0718**	C20H22O8	389.1241	0.5	-	-	-	+	Polydatin [a]
34	12.20	625.1434	-H	**301.0356**	C27H30O17	625.1410	3.8	+	+	+	-	Quercetin-3,4-diglucoside [a]
35	13.89	301.0002	-H	270.9953; 257.0102; 245.0096; **229.0152**	C14H6O8	300.9989	4.0	+	+	+	+	Ellagic acid [a]
36	13.97	463.0870	-H	300.0281; **301.0341**; 271.0245	C21H20O13	463.0882	-2.6	+	+	+	+	Quercetin-3-O-galactoside [a]
37	13.99	477.0680	-H	**301.0354**; 178.9969; 151.0025	C21H18O13	477.06746	1.1	+	+	+	+	Quercetin-3-O-glucoronide [a]
38	14.34	463.0890	-H	300.0281; **301.0341**; 271.0244	C21H20O12	463.0882	1.7	+	+	+	+	Quercetin-3-O-glucoside [a]
39	14.39	609.1483	-H	300.0287; **301.0351**	C27H30O16	609.14611	3.6	+	+	+	+	Quercetin-3-O-rutinoside [a]
40	14.68	433.1144	-H	**271.0606**	C21H22O10	433.114	0.9	+	+	+	+	Naringenin-7-O-glucoside [a]
41	15.61	461.0730	-H	447.0615; 315.0188; **285.0417**	C21H18O12	461.07255	1.0	+	+	+	+	Kaempferol-3-O-glucoronide
42	15.64	447.0572	-H	**315.0205**; 285.0418	C20H16O12	447.0569	0.7	+	+	+	+	Methylellagic acid pentose conjugate
43	15.64	435.1302	-H	**273.0774**; 229.0868; 167.0347	C21H24O10	435.12967	1.2	+	+	+	+	Phloridzin [a]
44	15.86	447.0942	-H	300.0281; **284.0326**	C21H20O11	447.09329	2.0	+	+	+	+	Kaempferol-3-O-glucoside [a]
45	15.97	475.0522	-H	432.0343; 329.1265; **300.9980**	C21H16O13	475.05181	0.8	+	+	-	+	Ellagic acid acetyl-pentose conjugate
46	17.73	301.0357	-H	**178.9970**; 151.0028; 121.0277	C15H10O7	301.03538	1.1	+	+	+	+	Quercetin [a]
47	19.01	273.0776	-H	**167.0357**; 125.0227; 123.0435; 119.0487	C15H14O5	273.07685	2.7	-	-	+	-	Phloretin [a]
48	26.37	503.3396	-H	**485.3325**; 441.485	C30H48O6	503.33781	3.6	+	+	-	+	Madecassic acid
49	27.22	487.3427	-H	**469.3341**; 425.3448	C30H48O5	487.3429	-0.4	+	+	-	-	Asiatic acid

Table 3 – Retention times (t_R, min), molecular and pseudo-molecular ions (TOF, Da), mass fragments (Q/TOF, Da), proposed formula, corresponding exact mass (Da) and accuracy (Δ, ppm) of peaks identified under positive ionization. Symbols + and − mean detected and not detected, respectively. GA, Glen Ample; T, Tulameen; J, Jewel; G, Fall Gold. a Peaks annotated at Level I. The base peak of each MS2 spectra is reported in bold character.

Peak	t_R	TOF	Charge	Q/TOF	Formula	Exact mass	Δ	GA	T	J	G	Tentative identification
50	7.07	611.1604		**287.0574**; 449.1067	C27H31O16	611.1621	-2.8	+	+	+	+	Cyanidin-3-O-sophoroside [a]
51	7.3	757.2191		757.1961; 611.1593; **287.0577**	C33H41O20	757.21912	0.1	+	+	+	-	Cyanidin-3-O-(2G-glucosylrutinoside)
52	7.38	449.1080		**287.062**	C21H21O11	449.10839	-0.8	+	+	+	+	Cyanidin-3-galactoside [a]
53	7.39	727.2073		581.1491; 433.1129; 281.0590	C32H39O19	727.20855	-1.7	+	+	+	-	Cyanidin 3-xylosylrutinoside
54	7.4	449.1085		**287.0603**	C21H21O11	449.1084	0.2	+	+	+	+	Cyanidin-3-O-glucoside [a]
55	7.44	481.0973	+H	**319.0462**	C21H20O13	481.0982	-1.9	-	-	-	+	Myricetin hexoside
56	7.57	595.1658		287.0626; **271.0657**	C27H31O15	595.1663	-0.9	+	+	+	+	Pelargonidin-3-O-sophoroside
57	7.64	595.1672		449.1072; **287.0689**	C27H31O15	595.1663	1.5	+	+	+	+	Cyanidin-3-O-rutinoside [a]
58	7.8	741.2243		549.1948; **271.0644**	C33H41O19	741.2242	0.1	+	+	-	-	Pelargonidin-3-O-(2G)-glucosylrutinoside
59	7.93	433.1135		305.1562; **271.0653**	C21H21O10	433.11347	0.1	+	+	+	-	Pelargonidin-3-O-glucoside
60	8.08	595.1660		449.1058; **287.0702**	C27H31O15	595.1663	-0.5	+	+	+	-	Cyanidin glycosyl-rhamnoside I
61	8.16	579.1693		453.0077; **271.0595**	C27H31O14	579.17138	-3.6	-	-	+	-	Pelargonidin-3-O-rutinoside
62	8.25	419.0971		**287.0543**	C20H19O10	419.09782	-1.7	+	-	+	-	Cyanidin-3-O-aldopentose
63	8.26	419.0975		301.0719; **287.0552**	C20H19O10	419.09782	-0.7	-	-	+	-	Cyanidin-3-arabinoside [a]
64	8.29	595.1662		449.1060; **287.0701**	C27H31O15	595.1663	-0.2	+	+	+	-	Cyanidin glycosyl-rhamnoside II
65	8.31	463.1251		**301.0731**	C22H23O11	463.12404	2.3	+	+	+	+	Peonidin-3-O-glucoside [a]
66	9.35	611.1608		**303.0517**	C27H31O16	611.16121	-0.7	-	-	+	-	Delphinidin-3-O-rutinoside
67	10.11	331.1545	+H	287.1258; 285.1118; **151.0736**; 137.0584	C19H22O5	331.1540	-0.3	+	+	+	+	Gibberellin A7
68	10.88	535.1080		487.2175; **287.5045**	C24H23O14	535.10878	-1.5	+	-	+	+	Cyanidin-3-O-malonyl-glucoside

Peak 3 evidenced a [M–H]– ion at *m/z* 633.07 and principal MS2 fragment at *m/z* 301.00 (i.e., ellagic acid), corresponding to the neutral loss of one galloyl-hexose unit. Another less intense dissociation product of the quasi-molecular ion was 331.07 (i.e., galloyl hexose), due to the loss of one hexahydroxydiphenoyl (HHDP) unit. This peak was therefore attributed to a galloyl-HHDP-hexose (Δ = 0.3 ppm), which has been previously reported in other fruit species [36,37] but not in raspberry. Peaks 7 and 19 showed double-charged ion with [M–2H]$^{2-}$/2 at *m/z* 783.07, corresponding to a MW of 1569 Da and mono-charged fragment ions at *m/z* 935.09, 933.07, 633.08, 617.04, 331.07 and 300.99. This fragmentation pattern was in accordance with the ones observed by Zhang et al. [23] and Mullen et al. [38] in raspberry and attributed to two isomers of sanguiin H-10. Peaks 22 and 23 also showed a double-charged pseudo-molecular ion at *m/z* 934.08, corresponding to a MW of 1869 Da. The fragmentation gave rise to mono-charged ions at *m/z* 915.06, consistent with the losses of a galloyl-di-HHDP-hexose (934 Da), H_2O and H_2; a further loss of H2O originated the fragment at *m/z* 897.05. The product ions at *m/z* 633.08 (galloyl-HHDP-hexose) and 301.01 (ellagic acid) were the result of the fragmentation of the pseudo-molecular ion in the mono-charged fragment at *m/z* 934.07 (not detected, probably due to the high abundance of the double-charged ion) and its successive cleavage. Accordingly, these peaks were attributed to sanguiin H-6 isomers, already annotated in *R. idaeus* [23,39]. Peak 8 showed a double-charged ion with [M–2H]$^{2-}$/2 at *m/z* 858.07, corresponding to a MW of 1717 Da. The fragmentation pattern included the product ions at *m/z* 935.09, 858.08, 633.08, 631.06, and 301.00, which are consistent with a sanguiin H-6 degalloylated, previously annotated by different authors in *R. idaeus* cultivars [23,39]. Peak 12 showed a mono-charged pseudo-molecular ion at *m/z* 633.08 which fragmented only in *m/z* 301.00. This peak was tentatively attributed (Level II) to corilagin, a low molecular weight ellagitannins previously identified in R. idaeus berries [23]. Peak 21 had a pseudo-molecular triple-charged ion at *m/z* 933.74, equivalent to a MW of 2805 Da. This peak fragmented in a double-charged ion at *m/z* 617.04 (corresponding to the loss of a di-HHDP-hexose-

galloyl-HHDP) and a mono-charged ellagic acid ion (*m/z* 301.00). Based on these findings, this peak was annotated as lambertianin C, elsewhere described in R. idaeus berries [23,39,40]. Peak 24 was characterized by a double-charged pseudo-molecular ion at *m/z* 551.04 (corresponding to a MW of 1104 Da) and fragments at *m/z* 469.01 (HHDP-hexose), 301.00 (ellagic acid) and 169.01 (gallic acid), thus being identified as a sanguiin H-2, in accordance with a previous annotation on *R. idaeus* fruits [23]. Peak 28 exhibited a double-charged pseudo-molecular ion at m/z 859.08, corresponding to a MW = 1721.2 Da and fragments at *m/z* 785.09, 633.08 and 301.00 (**Figure 1**). This fragmentation pattern is consistent with the occurrence of an ellagitannin-like nobotanin/malabathrin molecule (see **Scheme S3**), which is here annotated in raspberry for the first time.
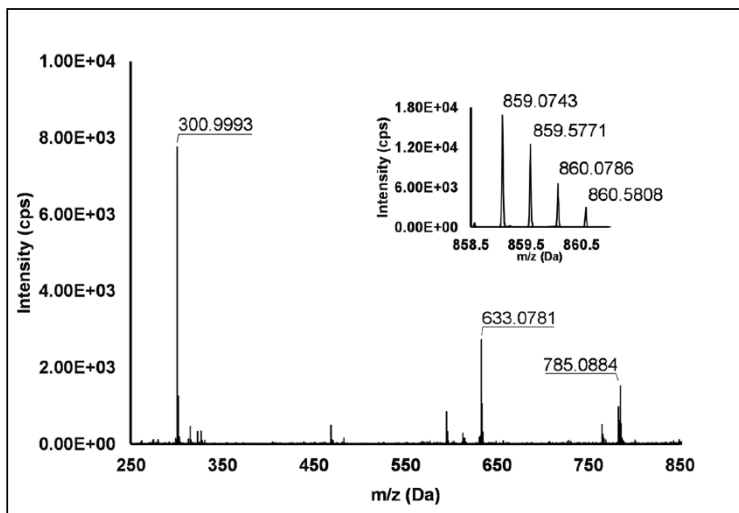

Figure 1 – TOF MS (top right) and Q/TOF MS2 spectra of Peak 28, tentatively identified as ellagitannin-like Nobotanin/Malabathrin.

Flavanols: Eight flavanols (Peaks 4, 5, 10, 11, 14, 16, 20 and 25) were identified under negative polarity in all genotypes, with the only exception of Peak 11, which was absent in *R. occidentalis*. Flavanols occurred mainly as proanthocyanidin dimers and trimers, whereas only (+)-catechin (Peak 10) and (−)-epicatechin (Peak 25) were annotated (Level I) as monomers. Peaks 4, 14 and 20 exhibited the same [M−H]⁻ quasi-molecular ion (*m/z* 577.14) and typical fragmentation of B-type (epi)catechin-(epi)catechin dimers, consisting in the retro-Diels–Alder

fission of the "C" ring (*m/z* 425.09), successive loss of water (*m/z* 407.08), the cleavage of the B-type linkage with formation of the (epi)catechin monomer (*m/z* 289.07) and the fission of the heterocyclic ring of the monomer (*m/z* 125.02) [1]. Peaks 14 and 20 were undoubtedly attributed to procyanidin B1 and procyanidin B2, respectively, based on the identity confirmation with authentic standards. Accordingly, Peak 4 was putatively ascribed to a B-type procyanidin isomer, in which the C4→C6 interflavanoid bond, instead of the C4→C8 one, is present between the two (epi)catechin units. Peaks 5 and 16 showed the same pseudo-molecular ion (*m/z* 863.20) and a fragmentation pattern consistent with the presence of monomer (*m/z* 287.06) and dimer (E)C units (*m/z* 577.13 and 575.12) derived from the quinone methide reaction, thus suggesting the attribution to B-type procyanidin trimers [41]. More in detail, for Peak 5, the undoubted attribution to procyanidin C1 was possible thanks to the availability of its authentic standard. Peak 11 showed a [M−H]⁻ pseudo-molecular ion at *m/z* 863.19 and a populated fragmentation pattern (i.e., m/z 711.14, 693.13, 575.12, 449.09, 423.07, 413.09 and 287.06) consistent with an A/B-linked procyanidin trimers (see **Figure 2**). In fact, as illustrated in Scheme S4, the product ions at *m/z* 711.14 derived from the retro Diels–Alder reaction affecting the B-type-linked (epi)catechin and gave rise to m/z 693.13 due to the loss of water. Moreover, *m/z* 287.06 and 575.12 were produced from the pseudo-molecular ion by the quinone methide reaction, the latter giving rise to *m/z* 449.09 and 423.08 (heterocyclic ring fissions). Flavanols in free and conjugated forms have been scarcely investigated in raspberry samples, since only Zhang and co-workers [23] investigated this compound class in depth, even though only one red raspberry variety was investigated and few metabolites were annotated. Hence, the present study adds new knowledge on the occurrence of flavanols in raspberry.
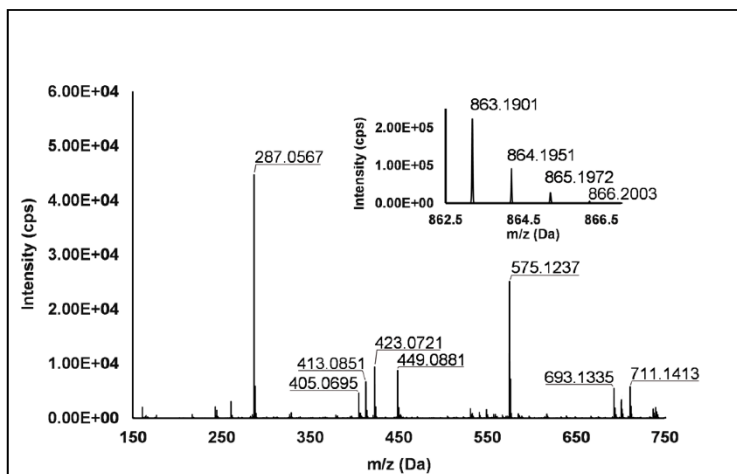
Figure 2 – TOF MS (top right) and Q/TOF MS2 spectra of Peak 11, tentatively identified as A/B type procyanidin trimer.

Flavonols: Nine flavonols were identified under negative polarity (**Table 1**), almost exclusively quercetin and kaempferol derivatives (Peaks 32, 34, 36–39, 41 and 44). Only one aglycone was found, i.e., quercetin (Peak 46), which was identified at Level I in all the investigated cultivars, thanks to its reference standard. Authentic standards allowed the Level I identification (**Table 1**) of quercetin-3-O-sophoroside (Peak 32) and quercetin-3,4-diglucoside (Peak 34) in *R. idaeus*, whereas quercetin-3-O-galactoside (Peak 36), quercetin-3-O-glucuronide (Peak 37), quercetin-3-O-glucoside (Peak 38), quercetin-3-O-rutinoside (Peak 39), kaempferol-3-O-glucuronide (Peak 41) and kaempferol-3-O-glucoside (Peak 44) were unequivocally annotated in all the investigated raspberry samples. These annotations have been described elsewhere in *R. idaeus* fruits [18,23,42], while they represent a first report in *R. occidentalis*. The MS2 fragmentation pattern of these features was characterized by the heterolytic and homolytic cleavages of the glycosidic bond that produced the characteristic $[Y_0]^-$ aglycone (i.e., *m/z* 301 for quercetin and m/z 285 for kaempferol) and/or $[Y0-H]^{\bullet-}$ (i.e., *m/z* 300 and 284) radical aglycone ions. Peak 13, identified only in J variety, exhibited a $[M-H]^-$ pseudo-molecular ion at *m/z* 609.15, which fragmented in *m/z* 301.04 and 300.03 (**Figure 3**), ascribable to the $[Y_0]^-$ ($\Delta = -1.0$ ppm) and $[Y0-H]^{\bullet-}$ ($\Delta = 3.7$ ppm) of quercetin aglycone. This fragmentation (see **Scheme**

S5) corresponds to the neutral loss of 308 Da, consistent with deoxyhexose-hexoside unit [1].
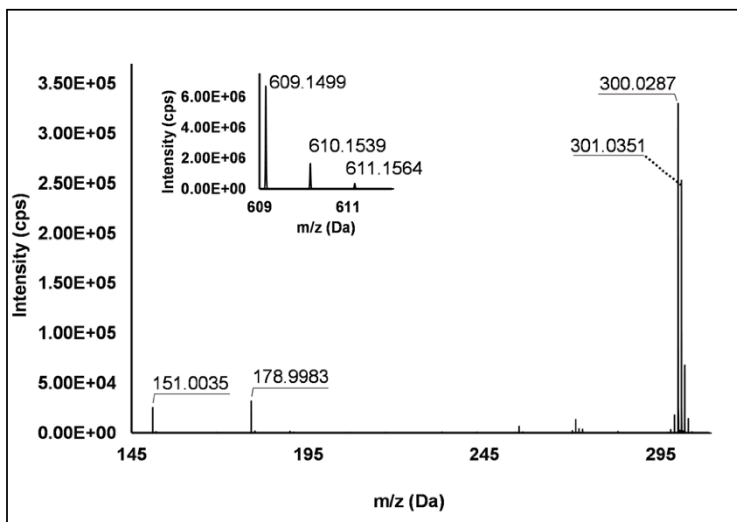

Figure 3 – TOF MS (top left) and Q/TOF MS2 spectrum of Peak 13, identified as quercetin-deoxyhexose-hexoside.

Further typical ions were observed at *m/z* 151.00 and *m/z* 179.00, due to different retrocyclization processes [43]. Thus, the feature was tentatively identified as quercetin-deoxyhexose-hexoside. Peak 15 exhibited a [M−H]⁻ pseudo-molecular ion at *m/z* 463.09, which is commonly found in many phenolic-rich fruits and attributed to flavonol glycosides [1,18]. However, the fragmentation pattern here observed highlighted product ions at *m/z* 327.05, 175.03 and 125.02, with no evidence of the aglycone ion formation (see **Figure 4**). Even though phenolic glycosides normally fragment through the cleavage of the glycosidic bond, the sugar moiety may undergo other bond breakings such as the cross-ring cleavage of the hexose part [44]. Based on the aforementioned findings, and considering the previously reported annotations of quercetin-3-O-hexosides (Peaks 36 and 38), this feature was tentatively attributed to a tetrahydroxyflavonol-3-O-hexoside with the aglycone different from quercetin (see **Scheme S6**). Peak 30 exhibited a [M−H]⁻ pseudo-molecular ion at *m/z* 651.20, which fragmented giving rise to a base peak at *m/z* 593.17 and main further ions at *m/z* 325.08, 285.04 and 284.03
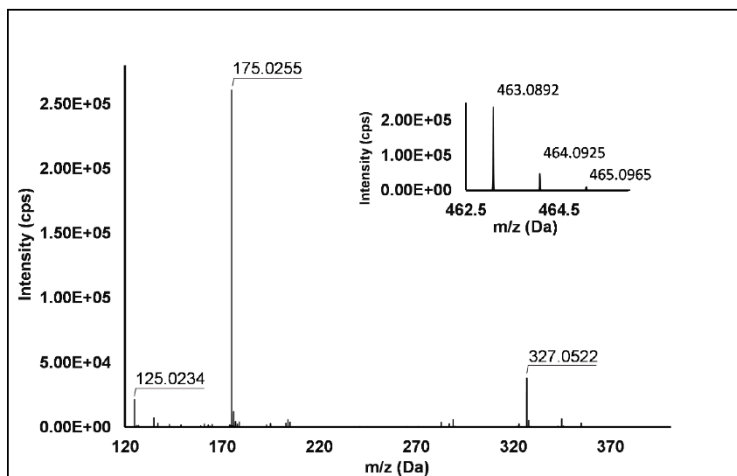
Figure 4 – TOF MS (top right) and Q/TOF MS2 spectra of Peak 15, tentatively identified as tetrahydroxyflavonol-3-O-hexoside.

(**Figure 5**). According to the hypothesized fragmentation illustrated in **Scheme S7**, these latter ions are consistent with a flavonoid scaffold and can be originated from the base peak by the formal loss of a deoxyhexose-hexoside moiety, due to the homolytic (*m/z* 284.03) or heterolytic (*m/z* 285.04) fission of the glycosidic bond. Finally, the product ion at *m/z* 325.08 derived from the cleavage of the carbon-oxygen bond on the aglycone ring. Thus, such structural information led to the tentative annotation of Peak 30 to a trihydroxy-methoxyflavone deoxyhexose-hexose derivative. One flavanol (Peak 55) was also identified under positive ionization (**Table 3**). In detail, this feature exhibited [M+H]$^+$ pseudo-molecular ion at *m/z* 481.10, which fragmented giving rise to the loss of 162.05 Da (hexose unit) and the product ion at *m/z* 319.05, consistent with the myricetin aglycone. The comparison with t_R and MS^2 fragmentation pattern of myricetin-3-O-glucoside reference standard confirmed the attribution to a myricetin hexoside but excluded the occurrence of the glucoside derivative. Myricetin derivatives have been identified as typical metabolomics traits of several berries [45], but this study revealed for the first time their occurrence in *R. idaeus* yellow variety.
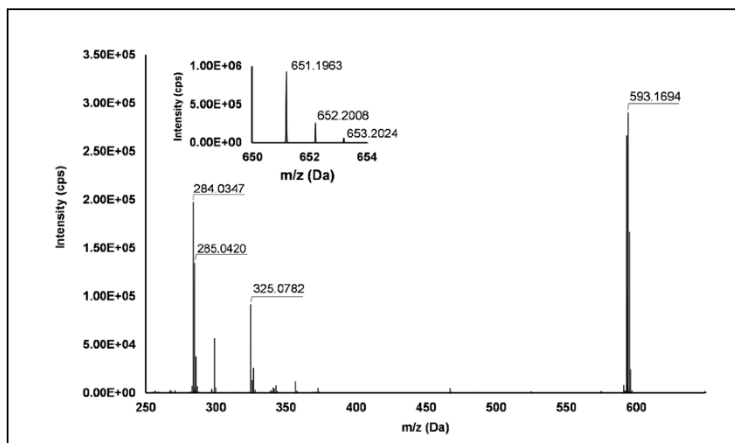
Figure 5 – TOF MS (top left) and Q/TOF MS2 spectra of Peak 30, tentatively identified as trihydroxy-methoxyflavone deoxyhexose-hexose derivative.

Anthocyanins: As illustrated in **Table 3**, the features identified in positive ionization mode were mostly anthocyanins, which were annotated at Levels I and II. In detail, the majority of the anthocyanins detected in the four raspberry varieties were cyanidin mono-, di- and tri-glycosides. These antioxidant compounds are well-known metabolic traits of raspberries, and in general of berries, being responsible for their red to purple pigmentation [19,46]. Peaks 50, 52, 54, 57 and 63 were annotated at Level I as cyanidin-3-O-sophoroside, cyanidin-3-O-galactoside, cyanidin-3-O-glucoside, cyanidin-3-O-rutinoside and cyanidin-3-O-arabinoside, respectively. These anthocyanins occurred in all the investigated raspberry fruit samples (intensity order: J >> T > GA > G), with the only exception of Peak 63, which was found only in R. occidentalis variety. Other cyanidin derivatives, elsewhere annotated in *Rubus* species [17,18,23,47], were identified here throughout the investigated samples. Among them, Peaks 53, 60 and 64 were annotated at Level II in GA, T and J cultivars, while they were absent in G variety. Peak 53 exhibited [M]$^+$ molecular ion at *m/z* 727.21 and MS^2 spectra highlighting successive losses of 146 Da (deoxyhexose unit) up to the formation of the well-known cyanidin aglycone at *m/z* 287.06. Thus, Peak 53 was putatively annotated as cyanidin-3-O-xylosilrutinoside, which was previously reported in *R. idaeus* berries [18]. Peaks 60 and 64 exhibited the same [M]$^+$ ion at *m/z* 595.17

and fragmentation pattern, which was characterized by successive losses of deoxyhexose (146 Da) and hexose (162 Da) units, giving rise to the cyanidin aglycone. These peaks were putatively annotated as two positional isomers of cyanidin diglycoside with hexose and rhamnose units. This attribution was consistent with: (i) their increased hydrophobicity compared to Peak 57 (cyanidin-3-O-rutinoside), ascribable to different linkages of the two glycosidic moieties (e.g., C3−glucose and C7−rhamnose) [48]; and (ii) the matches of their MS^2 spectra with that of Peak 57. Peak 62 showed a unique loss (132 Da) in its MS^2 spectra, leading to cyanidin aglycone fragment, which was ascribable to aldopentose loss [1]; moreover, thanks to the spectral matches with cyanidin-3-O-arabinoside reference standard, it was addressed at Level II as cyanidin-3-O-aldopentose. The remaining cyanidin derivatives were identified at Level II as cyanidin-3-O-(2G-glucosylrutinoside) (Peak 51, $\Delta = 0.1$ ppm) and cyanidin-3-O-malonyl-glucoside (Peak 68, $\Delta = -1.5$ ppm) with high mass accuracy, according to the isotopic profile of molecular ions and fragmentation pattern reported in previous studies [1,23]. Other fundamental anthocyanins in raspberries were pelargonidin derivatives, being their occurrence strongly dependent on fruit varieties [2], generally following the order J > T ≥ GA >> G. In detail, pelargonidin glycosides (**Table 3**) were putatively identified through the presence of the aglycone peak in the MS2 spectra (*m/z* 271.06) and of specific losses [18,23]: Peak 56 is the sophorose loss (324 Da,), Peak 58 is the successive losses of deoxyhexose (146 Da) and rutinose (324 Da) moieties, Peak 59 is the loss of glucose (162 Da) and Peak 61 is the deoxyhexose loss followed by glucose loss. Peak 56 occurred in the entire set of investigated raspberry cultivars, whereas Peaks 58 and 59 were both absent in yellow fruits, in which the anthocyanin content is strongly reduced [42]. As expected, pelargonidin-3-O-rutinoside was detected only in *R. occidentalis* black raspberries, since it has been already recognized as a specific trait of these fruits [2,16]. Thanks to the availability of reference standards, other two anthocyanins were identified at Level I in the four raspberry varieties; in detail, Peaks 65 and 66 were addressed as peonidin-3-O-

glucoside and delphinidin-3-O-rutinoside, respectively. The former compound was detected in all investigated samples, notably for the very first time in yellow cultivar, while the latter occurred only in *R. occidentalis* fruits. It should be noted that delphinidin derivatives have been generally associated to the secondary metabolism of bilberry, blueberry and blackcurrant [1,49], and they represent an uncommon trait for *R. idaeus* species, but they can occur in different raspberry species [50] as reported in this study. Overall, the occurrence profile of anthocyanins here presented (i.e., $J > T \geq GA > G$) agreed with previously reported data on raspberries. In fact, the content of anthocyanins in black cultivars is generally 3–10 times higher than in red fruits [46,51].

Other phenolics: Under negative ionization, nine further phenolic compounds (i.e., Peaks 26, 29, 31, 33, 35, 40, 42, 43, 45 and 47) were putatively or unequivocally identified in at least one of the four raspberry varieties investigated (**Table 2**). Peak 26 (detected only in G *Rubus idaeus* fruits) was characterized by a mono-charged precursor ion at m/z 341.12, which fragmented in m/z 179.07 and 121.03. The ion at m/z 341.12 derived from the loss of 162 Da (hexose unit) from the quasi-molecular ion and further fragmented in m/z 121.03 (loss of 57.03 Da). Interestingly, the two fragments are consistent with mass spectra of the available standard of coniferyl alcohol and literature findings [52]; accordingly, Peak 26 was undoubtedly identified as coniferin. This compound was previously annotated in the yellow raspberry cultivar "Heritage" and can therefore considered as a typical metabolic trait of yellow *R. idaeus* fruits [42]. Peaks 31, 35, 40, 43 and 47 were identified at Level I as taxifolin (flavanolol), ellagic acid, naringenin-7-O-glucoside (flavanone), phloridzin and phloretin (chalcones), confirming the annotations previously reported in raspberry varieties [23,25]. Peak 33 was also annotated at Level I and ascribed to polydatin (i.e., trihydroxystilbene-3-O-glucoside), the occurrence of which is conversely reported here for the first time in *R. occidentalis* berries. Peak 29 exhibited a [M–H]⁻ pseudo-molecular ion at *m/z* 593.15, and it was putatively annotated with good accuracy ($\Delta = -0.1$ ppm) as apigenin diglucoside in comparison with spectral libraries. In fact, its

fragmentation originated the product ions at *m/z* 475.14 (loss of 118 Da, retro-cyclization of the flavonoid C ring) and at *m/z* 431.06 (loss of 162 Da, consistent with a hexoside moiety). The latter further fragmented giving rise to the base peak at *m/z* 245.14, deriving from a retro-cyclization process and a consequent loss of 72 Da. Peaks 42 ([M−H]$^-$ = 447.06 Da) and 45 ([M−H]$^-$ = 475.05 Da) were preliminary recognized owing to the occurrence of fragments at *m/z* 315.02 and 301.00 (base peaks), which corresponded with good accuracy to methylellagic acid (Δ = 18.7 ppm) and ellagic acid (Δ = −3.3 ppm) units. The pseudo-molecular ion of Peak 42 loosed a pentoside group (i.e., 132 Da) giving rise to the aforementioned base peak, which further fragmented in m/z 285.04, consistently with the loss of a methoxy group (31 Da). Peak 42 was therefore putatively identified as methylellagic acid pentose conjugate. Instead, the precursor ion of Peak 45, besides base peak, was characterized by fragments at *m/z* 432.03 and 329.13, originated from a loss of 43 Da (acetyl moiety) and from the cross-ring cleavage of the saccharide moiety, respectively. Based on these considerations and on findings previously reported in the literature [53], this peak was tentatively annotated as ellagic acid acetyl-pentose conjugate.

Non-phenolic compounds: Among the features identified in negative ionization (**Table 2**), Peaks 48 ([M−H]$^-$ = 503.34 Da) and 49 ([M−H]$^-$ = 487.34 Da) were detected in at least two raspberry varieties and attributed to the free triterpenic acids (TTPAs) madecassic acid (Δ = 3.6 ppm) and asiatic acid (Δ = −0.4 ppm). The two peaks exhibited similar MS^2 fragmentation patterns, consistent to the aforementioned annotations, owing to successive neutral losses of water (18 Da) and CO_2 (44 Da). The isotopic profile of full scan MS and MS^2 spectra were also consistent with TTPAs [54]. The Level II identification of asiatic acid (Peaks 49) in T and GA varieties was in agreement with findings previously reported [55], while madecassic acid (Peak 48) was detected here for the first time in *R. idaeus* varieties. Among the set of features identified in positive ionization mode (**Table 3**), Peak 67 showed a [M + H]$^+$ pseudo-molecular ion at *m/z* 331.15 (Δ = −0.3 ppm), which occurred in all the raspberry sample here investigated. The

fragmentation pattern (**Figure 6**) was characterized by the cleavage of the five-carbon ring, which gave rise to *m/z* 151.07, and the successive loss of the methyl group producing *m/z* 137.06 (see **Scheme S8**). Accordingly, thanks to the analysis of available spectral libraries, Peak 67 was putatively ascribed to gibberellin A7. Although gibberellins have been already investigated as phytohormones responsible of flowering in *R. idaeus* plants [56,57,58], this study provided for the first time its occurrence in *R. occidentalis* J variety.
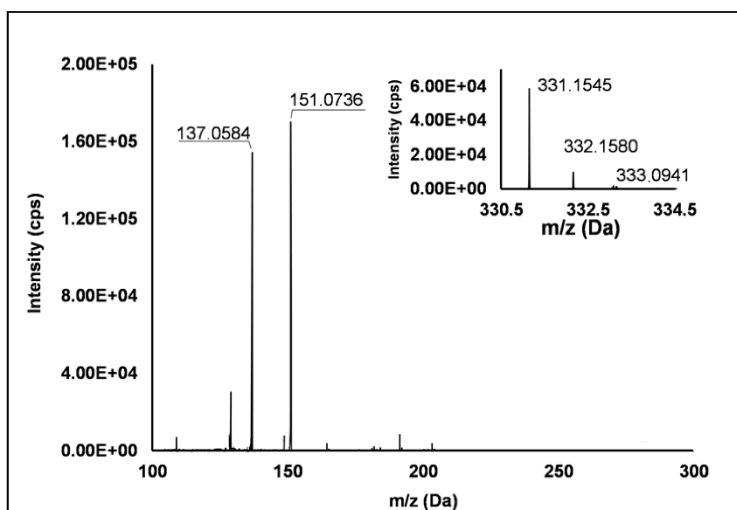
Figure 6 – TOF MS (top right) and Q/TOF MS2 spectra of Peak 67, tentatively identified as gibberellin A7.

3.3.3   Genotype Effects

A multivariate elaboration of the autoscaled original data (e.g., chromatographic feature intensities) was performed separately for compounds detected in negative and positive polarity, by means of PCA in order to discover possible clusterization among the investigated samples. As regards compounds detected in negative polarity, four principal components (PCs), characterized by eigenvalues > 1 and accounting for percentages of explained variances (EV%) of 55.4%, 19.5%, 13.1% and 5.2%, were obtained (total EV% = 93.2). Four significant PCs were also found for features identified in positive ionization, each one accounting for EV% of 55.2%, 16.6%, 13.6% and 8.4% (total EV% = 93.8%). **Figure 7**

summarizes the information obtained from the metabolomics profiling of the native compounds occurring in the investigated raspberry samples (including QCs), allowing for identifying more easily possible genotype effects: (i) among different cultivars of R. idaeus; and (ii) between R. idaeus and R. occidentalis species. In detail, **Figure 7A,B** illustrates the score plots of PC1 vs. PC2 of the significant features identified in negative and positive polarity, respectively, while **Figure 7C,D** shows the corresponding loading plots. In both datasets, the percentage of cumulative EV% of PC1 vs. PC2 space was > 70%, making the PCA results shown in **Figure 7** highly significant. Moreover, in both score plots, QCs were very close to the origin of the coordinates, indicating the high accuracy and precision of the entire analytical procedure. It should also be remarked that, in both score graphs, replicated samples showed very similar coordinates, thus indicating the homogeneous results obtained within each treatment. Both score plots showed a strong discrimination among different cultivars of *R. idaeus* and also between *R. idaeus* and *R. occidentalis* species, thus highlighting an effect of genotype on the expression of the annotated (poly)phenols. Interestingly, the four cultivars clustered according to their color not only as a function of their anthocyanin content (**Figure 7B**), but also in terms of the expression of other (poly)phenolics (**Figure 7A**). More in detail, as illustrated in **Figure 7C**, the red raspberries (i.e., GA and T) were mostly represented by ellagic acid and its derivatives and especially ellagitannins, which occurred in these cultivars with much higher intensities compared to J and G. Conversely, the G variety (*R. idaeus*) was mainly discriminated by the remarkable occurrence of flavanols, while the clusterization of the J cultivar (*R. occidentalis*) was due to the predominant presence of a miscellaneous compounds, which included mainly some flavonols and phenolic acids, one stilbene and two chalcones. A clearer scenario was highlighted for anthocyanins (**Figure 7D**), which were by far more abundant in J (*R. occidentalis*) as expected by its colorimetric analysis (**Table 1**). Cyanidin-3-O-sophoroside (Peak 50) and cyanidin-3-O-(2G-glucosylrutinoside) (Peak 51) were the only exceptions, being them more expressed in GA and T (R. idaeus).
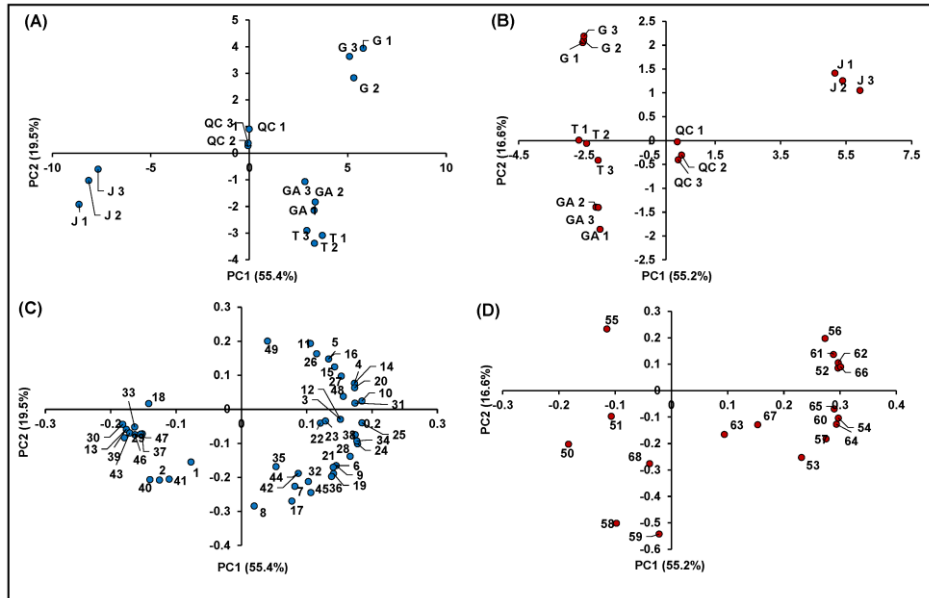
Figure 7 – Principal component analysis of the 68 identified significant features in *R. idaeus* and *R. occidentalis* fruits. (A) Score plot PC1 vs. PC2 of identified features in negative ionization; (B) score plot PC1 vs. PC2 of identified features in positive ionization; (C) loading plot PC1 vs. PC2 of identified features in negative ionization; and (D) loading plot PC1 vs. PC2 of identified features in positive ionization.

3.3.4   Correlations Between Identified Features and Pomological Parameters

Figure 8 illustrates the PPMC analysis of the features identified in both ionization modes in relation to the main pomological parameters (**Table 1**) of the investigated raspberries. A significant (p-value < 0.05) positive correlation was found between brightness (L) and the presence of free and condensed flavanols (i.e., Peaks 4, 5, 10, 16, 20 and 26), since these metabolites were found to be a distinctive metabolic trait of the yellow and brighter cultivar G, as observed in the PCA. A positive correlation, although characterized by a greater coefficient spread, was also observed between ellagitannins (i.e., Peaks 7, 8, 12, 19, 21–24 and 28) and some phenolic acid glycosides (i.e., Peaks 6, 9, 18 and 27). Moreover, the signal intensity of both these groups of compounds positively correlated with the "a" color coordinate, indicating that, as the fruit color turns red, the intensities of these (poly)phenols increased. Conversely, some anthocyanins exhibited an

unexpected statistically significant inverse correlation with respect to the coordinate "a". However, the predominant anthocyanins of raspberries, i.e., the Peaks 50, 51, 58 and 59, were strongly and significantly correlated with the red color of fruits. Interestingly, a positive correlation was found among anthocyanins, flavonols glycosides and TSS, suggesting that the expression of soluble sugars, which contribute extensively to this parameter, is related to the presence of some (poly)phenolic classes. As a final outcome, the acidity of raspberry fruits, expressed as titratable acids (TA), is significantly correlated to the presence of hydroxycinnamic acids (i.e., Peaks 6, 9 and 18), ellagitannins (i.e., Peaks 7, 8, 12, 19, 21–24 and 28) and ellagic acid and its derivatives (i.e., Peaks 35, 42 and 45).

3.4 Conclusions

This paper offers novel information on the (poly)phenolic composition of raspberry fruits, thanks to the adopted comprehensive untargeted strategy and data processing for features annotation, together with identity confirmation with authentic standards. Briefly, 68 bioactive compounds were successfully identified at Levels I and II, providing an in-depth characterization of the investigated raspberry genotypes. Even though most of the identified features belong to the already known major categories of (poly)phenols occurring in raspberry fruits (i.e., phenolic acids, ellagitannins, flavonols, flavanols and anthocyanins), this study extends the current knowledge of native composition on *R. idaeus* and above all *R. occidentalis*, the latter being scarcely investigated elsewhere. In fact, ferulic acid glycosides, one ellagitannin and madecassic acid were annotated for the first time in all the four investigated varieties. Additionally, the annotation of flavonol glycosides, polydatin and gibberellin A7 in *R. occidentalis* and of myricetin hexoside and peonidin-3-glucoside in the G raspberry cultivar of the *R. idaeus* species represent a first report in the investigated samples.
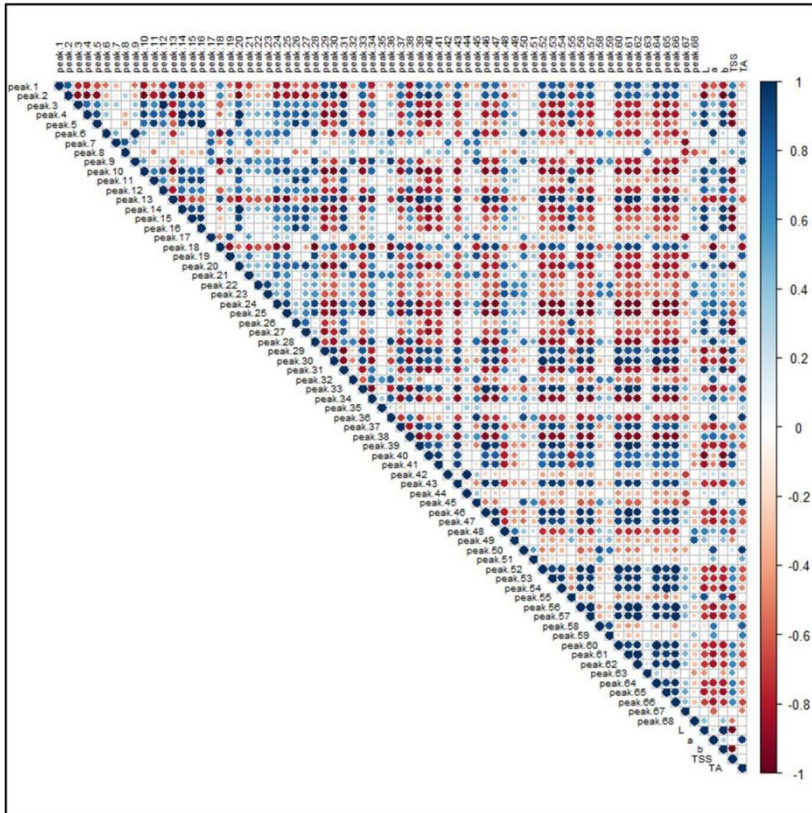
Figure 8 – Pearson product-moment correlation analysis of the 68 identified features in R. idaeus and R. occidentalis berries against pomological parameters. Features are indexed as reported in Table 1 (T1) and Table 2 (T2), respectively. Standard pomological data are reported in Table 1. L, brightness; a, red-shift; b, yellow-shift; TSS, total soluble solids (°Brix); TA, titratable acids (g citric acid kg−1 f.w.). The colored scale indicates the degree of correlation (white to blue color) or anti-correlation (white to red color).

The identification of few non-phenolic compounds possessing antioxidant properties, such as triterpenic acids, suggested the importance of extending untargeted investigations to compounds other than (poly)phenolic compounds. Overall, the in-depth profiling of raspberry secondary metabolites, together with the PCA carried out on the identified features, clearly highlighted the presence of genotype effects: (i) among different cultivars of *R. idaeus*; and (ii) between *R. idaeus* and *R. occidentalis* species. The latter represents an aspect investigated here for the first time.

Section 3

References

[1] C. Ancillotti, L. Ciofi, D. Rossini, U. Chiuminatto, J. Stahl-Zeng, S. Orlandini, S. Furlanetto, M. Del Bubba, Liquid chromatographic/electrospray ionization quadrupole/time of flight tandem mass spectrometric study of polyphenolic composition of different Vaccinium berry species and their comparative evaluation, Analytical and bioanalytical chemistry, 409 (2017) 1347-1368 https://doi.org/10.1007/s00216-016-0067-y.

[2] M. Kula, M. Majdan, D. Głód, M. Krauze-Baranowska, Phenolic composition of fruits from different cultivars of red and black raspberries grown in Poland, Journal of Food Composition and Analysis, 52 (2016) 74-82 https://doi.org/10.1016/j.jfca.2016.08.003.

[3] G. La Barbera, A.L. Capriotti, C. Cavaliere, S. Piovesana, R. Samperi, R.Z. Chiozzi, A. Laganà, Comprehensive polyphenol profiling of a strawberry extract (Fragaria× ananassa) by ultra-high-performance liquid chromatography coupled with high-resolution mass spectrometry, Analytical and bioanalytical chemistry, 409 (2017) 2127-2142 https://doi.org/10.1007/s00216-016-0159-8.

[4] B.M. Burton-Freeman, A.K. Sandhu, I. Edirisinghe, Red raspberries and their bioactive polyphenols: cardiometabolic and neuronal health links, Advances in Nutrition, 7 (2016) 44-65 https://doi.org/10.3945/an.115.009639.

[5] L. Mazzoni, P. Perez-Lopez, F. Giampieri, J.M. Alvarez-Suarez, M. Gasparrini, T.Y. Forbes-Hernandez, J.L. Quiles, B. Mezzetti, M. Battino, The genetic aspects of berries: from field to health, Journal of the Science of Food and Agriculture, 96 (2016) 365-371 https://doi.org/10.1002/jsfa.7216.

[6] M. Del Bubba, C. Di Serio, L. Renai, C.V.A. Scordo, L. Checchini, A. Ungar, F. Tarantini, R. Bartoletti, Vaccinium myrtillus L. extract and its native polyphenol-recombined mixture have anti-proliferative and pro-apoptotic effects on human prostate cancer cell lines, Phytotherapy Research, 35 (2020) 1089-1098 https://doi.org/10.1002/ptr.6879.

[7] V. Domazetovic, G. Marcucci, I. Falsetti, A.R. Bilia, M.T. Vincenzini, M.L. Brandi, T. Iantomasi, Blueberry Juice Antioxidants Protect Osteogenic Activity against Oxidative Stress and Improve Long-Term Activation of the Mineralization Process in Human Osteoblast-Like SaOS-2 Cells: Involvement of SIRT1, Antioxidants, 9 (2020) 125 https://doi.org/10.3390/antiox9020125.

[8] J. God, P.L. Tate, L.L. Larcom, Red raspberries have antioxidant effects that play a minor role in the killing of stomach and colon cancer cells, Nutrition Research, 30 (2010) 777-782 https://doi.org/10.1016/j.nutres.2010.10.004.

[9] A. Rodriguez-Mateos, R.D. Pino-García, T.W. George, A. Vidal-Diez, C. Heiss, J.P. Spencer, Impact of processing on the bioavailability and vascular effects of blueberry (poly) phenols, Molecular nutrition & food research, 58 (2014) 1952-1961 https://doi.org/10.1002/mnfr.201400231.

[10] E. Park, I. Edirisinghe, H. Wei, L.P. Vijayakumar, K. Banaszewski, J.C. Cappozzo, B. Burton-Freeman, A dose–response evaluation of freeze-dried strawberries independent of fiber content on metabolic indices in abdominally obese individuals with insulin resistance in a randomized, single-blinded, diet-

controlled crossover trial, Molecular nutrition & food research, 60 (2016) 1099-1109 https://doi.org/10.1002/mnfr.201500845.

[11] J. Lee, M. Dossett, C.E. Finn, Rubus fruit phenolic research: The good, the bad, and the confusing, Food Chemistry, 130 (2012) 785-796 https://doi.org/10.1016/j.foodchem.2011.08.022.

[12] G.-I. Hidalgo, M.P. Almajano, Red fruits: extraction of antioxidants, phenolic content, and radical scavenging determination: a review, Antioxidants, 6 (2017) 7 https://doi.org/10.3390/antiox6010007.

[13] D. Pinczinger, M.v. Reth, J. Keilwagen, T. Berner, A. Peil, H. Flachowsky, O.F. Emeriewen, Mapping of the Waxy Bloom Gene in 'Black Jewel'in a Parental Linkage Map of 'Black Jewel'×'Glen Ample'(Rubus) Interspecific Population, Agronomy, 10 (2020) 1579 https://doi.org/10.3390/agronomy10101579.

[14] FAOSTAT, Food and Agriculture Organization of United Nations, 2019.

[15] M.D. Teegarden, S.J. Schwartz, J.L. Cooperstone, Profiling the impact of thermal processing on black raspberry phytochemicals using untargeted metabolomics, Food chemistry, 274 (2019) 782-788 https://doi.org/10.1016/j.foodchem.2018.09.053.

[16] L. Paudel, F.J. Wyzgoski, M.M. Giusti, J.L. Johnson, P.L. Rinaldi, J.C. Scheerens, A.M. Chanon, J.A. Bomser, A.R. Miller, J.K. Hardy, NMR-based metabolomic investigation of bioactivity of chemical constituents in black raspberry (Rubus occidentalis L.) fruit extracts, Journal of agricultural and food chemistry, 62 (2014) 1989-1998 https://doi.org/10.1021/jf404998k.

[17] S.P. Mazur, A. Nes, A.-B. Wold, S.F. Remberg, K. Aaby, Quality and chemical composition of ten red raspberry (Rubus idaeus L.) genotypes during three harvest seasons, Food chemistry, 160 (2014) 233-240 https://doi.org/10.1016/j.foodchem.2014.02.174.

[18] I. Dincheva, I. Badjakov, V. Kondakova, P. Dobson, G. Mcdougall, D. Stewart, Identification of the phenolic components in Bulgarian raspberry cultivars by LC-ESI-MSn, Int. J. Agric. Sci. Res, 3 (2013) 137-138

[19] W. Mullen, M.E. Lean, A. Crozier, Rapid characterization of anthocyanins in red raspberry fruit by high-performance liquid chromatography coupled to single quadrupole mass spectrometry, Journal of Chromatography A, 966 (2002) 63-70 https://doi.org/10.1016/S0021-9673(02)00699-4.

[20] S.F. Remberg, A. Sønsteby, K. Aaby, O.M. Heide, Influence of postflowering temperature on fruit size and chemical composition of Glen Ample raspberry (Rubus idaeus L.), Journal of agricultural and food chemistry, 58 (2010) 9120-9128 https://doi.org/10.1021/jf101736q.

[21] G. Borges, A. Degeneve, W. Mullen, A. Crozier, Identification of flavonoid and phenolic antioxidants in black currants, blueberries, raspberries, red currants, and cranberries, Journal of agricultural and food chemistry, 58 (2009) 3901-3909 https://doi.org/10.1021/jf902263n.

[22] C.M. Bradish, P. Perkins-Veazie, G.E. Fernandez, G. Xie, W. Jia, Comparison of flavonoid composition of red raspberries (Rubus idaeus L.) grown in the Southern United States, Journal of agricultural and food chemistry, 60 (2011) 5779-5786 https://doi.org/10.1021/jf203474e.

[23] X. Zhang, A. Sandhu, I. Edirisinghe, B. Burton-Freeman, An exploratory study of red raspberry (Rubus idaeus L.)(poly) phenols/metabolites in human biological samples, Food & function, 9 (2018) 806-818 https://doi.org/10.1039/C7FO00893G.

[24] C. Ancillotti, M. Ulaszewska, F. Mattivi, M. Del Bubba, Untargeted Metabolomics Analytical Strategy Based on Liquid Chromatography/Electrospray Ionization Linear Ion Trap Quadrupole/Orbitrap Mass Spectrometry for Discovering New Polyphenol Metabolites in Human Biofluids after Acute Ingestion of Vaccinium myrtillus Berry Supplement, Journal of The American Society for Mass Spectrometry, 30 (2019) 381-402 https://doi.org/10.1007/s13361-018-2111-y.

[25] E. Carvalho, P. Franceschi, A. Feller, L. Palmieri, R. Wehrens, S. Martens, A targeted metabolomics approach to understand differences in flavonoid biosynthesis in red and yellow raspberries, Plant Physiology and Biochemistry, 72 (2013) 79-86 https://doi.org/10.1016/j.plaphy.2013.04.001.

[26] L. Renai, F. Tozzi, C.V. Scordo, E. Giordani, M.C. Bruzzoniti, D. Fibbi, L. Mandi, N. Ouazzani, M. Del Bubba, Productivity and nutritional and nutraceutical value of strawberry fruits (Fragaria x ananassa Duch.) cultivated under irrigation with treated wastewaters, Journal of the Science of Food and Agriculture, 101 (2021) 1239-1246 https://doi.org/10.1002/jsfa.10737.

[27] M. Del Bubba, E. Giordani, C. Ancillotti, W.A. Petrucci, L. Ciofi, D. Morelli, C. Marinelli, L. Checchini, S. Furlanetto, Morphological, nutraceutical and sensorial properties of cultivated Fragaria vesca L. berries: influence of genotype, plant age, fertilization treatment on the overall fruit quality, Agricultural and Food Science, 25 (2016) 187-201 https://doi.org/10.23986/afsci.56867.

[28] F. Tozzi, P. Legua, J.J. Martínez-Nicolás, D. Núñez-Gómez, E. Giordani, P. Melgarejo, Morphological and nutraceutical characterization of six pomegranate cultivars of global commercial interest, Scientia Horticulturae, 272 (2020) 109557 https://doi.org/10.1016/j.scienta.2020.109557.

[29] C.E. Douglas, F.A. Michael, On distribution-free multiple comparisons in the one-way analysis of variance, Communications in Statistics-Theory and Methods, 20 (1991) 127-139 https://doi.org/10.1080/03610929108830487.

[30] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, Proposed minimum reporting standards for chemical analysis, Metabolomics, 3 (2007) 211-221 https://doi.org/10.1007/s11306-007-0082-2.

[31] E. Krüger, H. Dietrich, E. Schöpplein, S. Rasim, P. Kürbel, Cultivar, storage conditions and ripening effects on physical and chemical qualities of red raspberry fruit, Postharvest Biology and Technology, 60 (2011) 31-37 https://doi.org/10.1016/j.postharvbio.2010.12.001.

[32] G.E. Pantelidis, M. Vasilakakis, G.A. Manganaris, G. Diamantidis, Antioxidant capacity, phenol, anthocyanin and ascorbic acid contents in raspberries, blackberries, red currants, gooseberries and Cornelian cherries, Food chemistry, 102 (2007) 777-783 https://doi.org/10.1016/j.foodchem.2006.06.021.

[33] J.A. Stavang, S. Freitag, A. Foito, S. Verrall, O.M. Heide, D. Stewart, A. Sønsteby, Raspberry fruit quality changes during ripening and storage as assessed by colour, sensory evaluation and chemical analyses, Scientia Horticulturae, 195 (2015) 216-225 https://doi.org/10.1016/j.scienta.2015.08.045.

[34] S.Y. Wang, C.-T. Chen, C.Y. Wang, The influence of light and maturity on fruit quality and flavonoid content of red raspberries, Food chemistry, 112 (2009) 676-684 https://doi.org/10.1016/j.foodchem.2008.06.032.

[35] M. Schulz, J.F. Chim, Nutritional and bioactive value of Rubus berries, Food Bioscience, 31 (2019) 100438 https://doi.org/10.1016/j.fbio.2019.100438.

[36] M. Del Bubba, L. Checchini, U. Chiuminatto, S. Doumett, D. Fibbi, E. Giordani, Liquid chromatographic/electrospray ionization tandem mass spectrometric study of polyphenolic composition of four cultivars of Fragaria vesca L. berries and their comparative evaluation, Journal of Mass Spectrometry, 47 (2012) 1207-1220 https://doi.org/10.1002/jms.3030.

[37] T.J. Hager, L.R. Howard, R. Liyanage, J.O. Lay, R.L. Prior, Ellagitannin composition of blackberry as determined by HPLC-ESI-MS and MALDI-TOF-MS, Journal of Agricultural and Food Chemistry, 56 (2008) 661-669 https://doi.org/10.1021/jf071990b.

[38] W. Mullen, T. Yokota, M.E. Lean, A. Crozier, Analysis of ellagitannins and conjugates of ellagic acid and quercetin in raspberry fruits by LC–MSn, Phytochemistry, 64 (2003) 617-624 https://doi.org/10.1016/S0031-9422(03)00281-4.

[39] G. McDougall, I. Martinussen, D. Stewart, Towards fruitful metabolomics: high throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry, Journal of Chromatography B, 871 (2008) 362-369 https://doi.org/10.1016/j.jchromb.2008.06.032.

[40] W. Mullen, J. McGinn, M.E. Lean, M.R. MacLean, P. Gardner, G.G. Duthie, T. Yokota, A. Crozier, Ellagitannins, flavonoids, and other phenolics in red raspberries and their contribution to antioxidant capacity and vasorelaxation properties, Journal of Agricultural and Food Chemistry, 50 (2002) 5191-5196 https://doi.org/10.1021/jf020140n.

[41] N. Teixeira, J. Azevedo, N. Mateus, V. de Freitas, Proanthocyanidin screening by LC–ESI-MS of Portuguese red wines made with teinturier grapes, Food chemistry, 190 (2016) 300-307 https://doi.org/10.1016/j.foodchem.2015.05.065.

[42] E. Carvalho, P.D. Fraser, S. Martens, Carotenoids and tocopherols in yellow and red raspberries, Food chemistry, 139 (2013) 744-752 https://doi.org/10.1016/j.foodchem.2012.12.047.

[43] G.C. Justino, C.M. Borges, M.H. Florêncio, Electrospray ionization tandem mass spectrometry fragmentation of protonated flavone and flavonol aglycones: a re-examination, Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 23 (2009) 237-248 https://doi.org/10.1002/rcm.3869.

[44] T. Yuan, X.-F. Guo, S.-Y. Shao, R.-M. An, J. Wang, J. Sun, Characterization and identification of flavonoids from Bambusa chungii leaves extract by UPLC-

ESI-Q-TOF-MS/MS, Acta Chromatographica, 33 (2020) 281-294 https://doi.org/10.1556/1326.2020.00777.

[45] M. Mikulic-Petkovsek, A. Slatnar, F. Stampar, R. Veberic, HPLC–MSn identification and quantification of flavonol glycosides in 28 wild and cultivated berry species, Food Chemistry, 135 (2012) 2138-2146 https://doi.org/10.1016/j.foodchem.2012.06.115.

[46] R. Bobinaitė, P. Viškelis, P.R. Venskutonis, Variation of total phenolics, anthocyanins, ellagic acid and radical scavenging capacity in various raspberry (Rubus spp.) cultivars, Food Chemistry, 132 (2012) 1495-1501 https://doi.org/10.1016/j.foodchem.2011.11.137.

[47] L. Kaume, L.R. Howard, L. Devareddy, The blackberry fruit: a review on its composition and chemistry, metabolism and bioavailability, and health benefits, Journal of agricultural and food chemistry, 60 (2011) 5716-5727 https://doi.org/10.1021/jf203318p.

[48] I. Regos, A. Urbanella, D. Treutter, Identification and quantification of phenolic compounds from the forage legume sainfoin (Onobrychis viciifolia), Journal of Agricultural and Food Chemistry, 57 (2009) 5843-5852 https://doi.org/10.1021/jf900625r.

[49] L. Chen, X. Xin, Q. Yuan, D. Su, W. Liu, Phytochemical properties and antioxidant capacities of various colored berries, Journal of the Science of Food and Agriculture, 94 (2014) 180-188 https://doi.org/10.1002/jsfa.6216.

[50] Y. Fu, X. Zhou, S. Chen, Y. Sun, Y. Shen, X. Ye, Chemical composition and antioxidant activity of Chinese wild raspberry (Rubus hirsutus Thunb.), LWT-Food Science and Technology, 60 (2015) 1262-1268 https://doi.org/10.1016/j.lwt.2014.09.002.

[51] C. Weber, P. Perkins-Veazie, P. Moore, L. Howard, Variability of antioxidant content in raspberry germplasm, IX International Rubus and Ribes Symposium 777, 2005, pp. 493-498 https://doi.org/10.17660/ActaHortic.2008.777.75.

[52] G. Tóth, C. Barabás, A. Tóth, Á. Kéry, S. Béni, I. Boldizsár, E. Varga, B. Noszál, Characterization of antioxidant phenolics in Syringa vulgaris L. flowers and fruits by HPLC-DAD-ESI-MS, Biomedical chromatography, 30 (2016) 923-932 https://doi.org/10.1002/bmc.3630.

[53] N. de Andrade Neves, P.C. Stringheta, S. Gómez-Alonso, I. Hermosín-Gutiérrez, Flavonols and ellagic acid derivatives in peels of different species of jabuticaba (Plinia spp.) identified by HPLC-DAD-ESI/MSn, Food chemistry, 252 (2018) 61-71 https://doi.org/10.1016/j.foodchem.2018.01.078.

[54] S. Guo, J.-A. Duan, Y.-P. Tang, N.-Y. Yang, D.-W. Qian, S.-L. Su, E.-X. Shang, Characterization of triterpenic acids in fruits of Ziziphus species by HPLC-ELSD-MS, Journal of agricultural and food chemistry, 58 (2010) 6285-6289 https://doi.org/10.1016/j.lwt.2014.09.002.

[55] Y. Wang, Y. Suo, Y. Sun, J. You, Determination of triterpene acids from 37 different varieties of raspberry using pre-column derivatization and HPLC fluorescence detection, Chromatographia, 79 (2016) 1515-1525 https://doi.org/10.1007/s10337-016-3174-1.

[56] L. Chai, Y. Li, S. Chen, A. Perl, F. Zhao, H. Ma, RNA sequencing reveals high resolution expression change of major plant hormone pathway genes after young seedless grape berries treated with gibberellin, Plant Science, 229 (2014) 215-224 https://doi.org/10.1016/j.plantsci.2014.09.010.

[57] F.J. Pérez, C. Viani, J. Retamales, Bioactive gibberellins in seeded and seedless grapes: identification and changes in content during berry development, American journal of enology and viticulture, 51 (2000) https://doi.org/315-318 10.5344/ajev.2000.51.4.315.

[58] P. Palonen, E. Pehkonen, M. Rantanen, Growth control of "Glen Ample" and "Tulameen" raspberry cultivars with single and repeated ProCa applications, Eur J Hortic Sci, 78 (2013) 22-29.

4 Quality by design optimization of a liquid chromatographic-tandem mass spectrometric method for the simultaneous analysis of structurally heterogeneous pharmaceutical compounds and its application to the rapid screening in wastewater and surface water samples by large volume direct injection.

Abstract

This study focused on the Analytical Quality by Design (AQbD) optimization of the chromatographic separation and mass spectrometric detection of a wide group of structurally heterogeneous model pharmaceutical compounds (PhCs) and transformation products (TPs), chosen to cover the challenging issues of the co-presence of compounds characterized by (i) a wide range of physicochemical properties, (ii) the same mass transitions, and (iii) different ionisation modes. Italian consumption of PhCs were also considered as election criteria of target analytes. Octadecyl and pentafluorophenyl stationary phases, acetonitrile/methanol ratios and acidity of the eluents, column temperature, initial organic phase percentage, and elution gradient were investigated by AQbD, aiming at optimizing critical resolutions, sensitivities, and analysis time. Statistically significant models were obtained in most cases with fitting and cross-validation coefficients in the ranges of 0.681-0.998 and 0.514-0.967, respectively. After optimization, the analysis of target analytes was performed in a single chromatographic run, adopting a mixed acquisition mode based on scheduled acquisition windows comprising both single polarity and continuous polarity switching. For most investigated analytes the method provided detection limits in the sub-ng L^{-1} to low ng L^{-1} range, meeting for macrolides the sensitivity requested by the "Watch List" 2018/840/EU. The optimized method was applied to the direct injection analysis of PhCs and TPs in four wastewater treatment plant (WWTP) effluents and surface water (SW) samples collected in the receiving water bodies. Absolute values of matrix effect were found to be far higher than 20% for most target analytes in most samples. Seventeen PhCs and two TPs were quantified in at least one sample, at the wide concentration range of about 1-3200 ng L^{-1}. The most occurring PhCs in both WWTP effluents and SWs were levofloxacin (202-1239 and 100-830 ng L^{-1}), furosemide (865-3234 and 230-880 ng L^{-1}), ketoprofen (295-1104 and 270-490 ng L^{-1}), and ibuprofen (886-3232 and 690-1440 ng L^{-1}).

4.1 Introduction

Modern environmental analytical chemistry constantly deals with the development of innovative, multiresidue and high-throughput methods for quantifying contaminants of emerging concern (CECs) in environmental matrices [1]. In recent years, among the several classes of CECs, pharmaceutical compounds (PhCs) have caught the attention of the scientific community for their wide environmental diffusion [2] and possible negative effects towards living organisms, including humans. In fact, PhCs are characterized by a high biological activity, which results in a number of potential adverse effects when they are released in the environment, such as the endocrine disrupting activity exerted by estrogens or the induction of antibiotic resistance phenomena in human pathogenic and/or opportunistic microorganisms [3, 4]. Human consumption, animal farms, and pharmaceutical industries actively contribute to the presence of PhCs in the environment, which reach surface waters through wastewater treatment plants (WWTPs) [5]. Indeed, several classes of PhCs have been frequently detected in WWTP effluents at ng L^{-1} to µg L^{-1} levels, especially those dealing with urban wastewaters [2, 6]. In fact, these WWTPs are in most cases not specifically designed for PhCs removal [7], thus representing a point source for their release in the environment [8]. Among these classes, the most frequently detected are antibiotics (e.g. amoxicillin and clarithromycin) [2, 6], non-steroidal anti-inflammatory drugs (e.g. diclofenac and ibuprofen) [9], beta-blockers (e.g. atenolol) [7], antiepileptics (e.g. carbamazepine) [8], estrogens (e.g. estrone and 17-β-estradiol) [10], and psycholeptics (e.g. diazepam) [2, 7]. The quantitative determination of PhCs in wastewaters and environmental waters is in most cases performed by liquid chromatography hyphenated with electrospray ionization (ESI) tandem mass spectrometry (LC-MS/MS), mainly adopting the targeted detection mode and using triple quadrupole as mass detector [11], which provides the highest instrumental sensitivity, compared to other mass detection approaches

[12]. However, it should be underlined that PhCs have a high structural variability, since they include molecules with very different molecular weights and acid-base properties, thus resulting in different mass ionization characteristics and chromatographic behaviours. Moreover, the environmental monitoring of PhCs often involves the analysis of isobaric compounds with the same MS/MS transitions, for which the baseline chromatographic separation is mandatory. Accordingly, the development of multiresidue LC-MS/MS methods for the determination of PhCs in the environment undoubtedly represents a complex and challenging issue in analytical chemistry. Based on literature, this issue is commonly solved by adopting two distinct chromatographic runs for analytes detectable under positive ionisation (PI) and negative ionisation (NI) modes [13, 14], with an obvious negative impact in terms of analytical throughput. Alternatively, the continuous NI/PI polarity switching (CPS) on the entire chromatographic run is applied [15, 16], which may however causes a loss in sensitivity, accuracy, and precision, due to classical source phenomena that affect the ionisation process [17]. The acquisition problems of polarity switching are even more accentuated when it is applied on numerous transitions, as in the case of its use along the entire chromatographic run of a multiresidue method. To overcome these problems, an acquisition method using the CPS along the entire chromatographic run and providing the specific setting of dwell time for each acquired transition, has been recently proposed for the simultaneous PI/NI determination of 23 PhCs and 20 transformation products [18]. However, to our knowledge, the optimization of chromatographic runs for the multiresidue analysis of PhCs has not been investigated in order to limit and/or eliminate the use of CPS and the above-mentioned related disadvantages.

Based on these considerations, this study focused on the optimization of the chromatographic separation and scheduled MRM detection of a large number of PhCs, without using CPS throughout the whole chromatogram, aiming at solving the issues of co-presence of compounds characterized by (i) a wide range of physicochemical properties (e.g. large variations in molecular weight and acid-

base characteristics), (ii) the same mass transitions, and (iii) different ionisation modes. For this purpose, a model group of 34 PhCs and related transformation products (TPs) was selected, considering the aforementioned issues. Italian consumption data of PhCs in the latest years were also considered as election criteria [19]. Moreover, all PhCs reported in the most recent European "Watch List" for surface water monitoring [20] were included in this study, regardless their consumption data. The optimization of the chromatographic method was carried out adopting the Analytical Quality by Design (AQbD) approach, which has been applied here by using several chromatographic input variables for modelling numerous resolutions and ESI-MS response factors, i.e. realizing an innovative and complex application of AQbD. Moreover, even though AQbD concepts have been successfully applied to the determination of active ingredients and impurities in medicinal products [21-24] and the analysis of food constituents [25, 26], there is no AQbD workflow applied to the LC-MS/MS analysis of PhCs in an environmental context, especially taking into account the high complexity in terms of input and output variables here investigated. The optimized LC-MS/MS method was applied to the large volume direct injection (LVDI) analysis of treated wastewater (TWW) and surface water (SW) samples, as a tool for the rapid screening of CECs in aqueous samples [27], evaluating the matrix effect (ME) associated to the method developed.

4.2   Materials and methods

4.2.1   Solvents and materials

Formic acid (HCOOH, MS grade, purity ≥ 99.0%) and dimethyl sulfoxide (DMSO, purity ≥ 99.9%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water, methanol (MeOH) and acetonitrile (ACN) were obtained from Carlo Erba (Milan, Italy). Ultrapure water (resistivity > 18 MΩ) was obtained from a Milli-Q system (Millipore, Billerica, MA, USA). Disodium EDTA (purity > 99%) was purchased from Phargam s.r.l. (Milan, Italy). Minisart® regenerated cellulose (RC) syringe filters were supplied by Sartorius (Gottingen,

Germany). The chromatographic columns used in this study were the endcapped pellicular Kinetex® C18 and Kinetex® pentafluorophenyl (PFP) columns (Phenomenex, Torrance, CA, USA), both 100x3 mm, particle size 2.6 μm.

4.2.2   Target analytes

Target compounds and internal standards of this study (Sigma-Aldrich) included the analgesics, antibiotics, beta-blockers, urinary alpha-blockers, psychiatric drugs, diuretics, statins, gastric protectors and estrogens reported in **Table S1** and **Figure S1** (Section 1 of the *Supplementary material*).

4.2.3   Preparation of standard stock and working solutions

Standard stock solutions of individual PhCs (1 mg mL^{-1}) were prepared by weighing each solid standard and dissolving it in the appropriate solvent. MeOH or ACN resulted suitable for the majority of the selected analytes, whereas AMX and CIP were respectively dissolved in DMSO and in hydrochloric acid 0.1 M in LC-MS grade water. Due to the different instrumental responses and chromatographic behaviour of each analyte, evaluated by injections of individual standard compounds, a multi-compound standard mixture ("solution 1") was reconstructed in LC-MS grade water at the concentration level of each PhC to obtain a signal-to-noise ratio (S/N) equal to 1000 (see **Table S1**). Finally, the working standard solution ("solution 2") used for the optimization procedure was obtained by diluting "solution 1" with a 4 mM EDTA in LC-MS grade water [28] at the 1/1 (v/v) ratio (S/N ~ 500). An EDTA solution in LC-MS grade water at the same concentration of "solution 1" (i.e. 2 mM) was used as procedural blank.

4.2.4   Water samples

Effluent wastewater samples were collected from two WWTPs treating mixed industrial-municipal sewages (TWW1 e TWW2), and two WWTPs receiving municipal wastewaters including hospital sewages (TWW3 and TWW4). SW samples (SW1, SW2, SW3 and SW4) were collected in the receiving water bodies within one km distance from the WWTP discharge. These samples were analysed

for obtaining information on the extent of ME, which is commonly a drawback when LVDI is adopted. All samples were pre-treated by adding EDTA as chelating agent to prevent the formation of β-lactams epimers and successive transformation products that are catalysed by the presence of metal ions [29]. Sampling location coordinates (**Table S2**), characteristics of WWTPs, and details of sampling procedure and sample pretreatment are reported in Section 2 of the *Supplementary material*.

4.2.5   LC-MS/MS analysis

Instrumental analysis was performed on a Shimadzu (Kyoto, Japan) chromatographic system coupled with a 5500 QTrap mass spectrometer (Sciex, Framingham, MA, USA), equipped with a Turbo V® interface by an ESI probe (see Section 3 of the *Supplementary material* for full details of the instrumentation). Source and compound dependent parameters (**Table S3**) were optimized and set as described in Section 3 of the *Supplementary material*. MS/MS analysis was carried out using a polarity-switched and time-scheduled Multiple Reaction Monitoring (MRM) in NI (ESI (-)) and PI (ESI (+)). The optimized chromatographic conditions are summarised hereafter. Chromatographic column: Kinetex® PFP; eluent "A": acidic water (0.011% HCOOH); eluent "B": ACN/MeOH ratio 1.75 (v/v); flow rate: 0.50 mL min^{-1}; temperature: 32 °C; elution gradient: initial isocratic at "B" percentage 11.0 % for 1 min and increase of the eluent "B" at 11.5% per minute for 8.7 min. The injection volume was 100 µL. In order to minimize MS source contamination, from 0 min to 1 min and from 10 min to the end of the chromatographic run, the LC eluate was diverted to waste by means of a two-position six-port valve (model HT, Vici, Schenkon, Switzerland) installed upstream the mass spectrometer.

4.2.6   AQbD workflow

AQbD focuses on the definition of the method operable design region (MODR), which allows to identify the multidimensional region of the experimental domain, defined by the "critical method parameters" (CMPs), where the analytical

performances meet the selected "critical method attributes" (CMAs), maintaining the risk of error below the desired threshold [30]. The probability of meeting specifications for all the CMAs is computed by the simultaneous use of calculated models by experimental design and of Monte Carlo simulations [31], enabling the definition of the MODR. The investigation of the knowledge space and the MODR identification were performed adopting NemrodW software (NemrodW sarl, France) for setting up the asymmetric screening matrix and the Plackett-Burman designs. MODDE 12.1 software (MKS Umetrics AB, Sweden) was employed for planning the Central Composite Orthogonal Design (CCO) in Response Surface Methodology (RSM), for the related data treatment and MODR identification.

4.2.7 *Figures of merit of the method for LVDI analysis of PhCs and TPs*

The performances of the optimized LVDI method were evaluated through method detection (MDL) and quantification (MQL) limits, intra- and inter-day precision, and linearity range. MDLs and MQLs were assessed by replicated (n=10) analysis of procedural blank, according to the following equation [27]:

$$MDL\ (MQL) = (k \cdot \sigma b)/m \qquad (1)$$

where k is the critical value given by 2t1-α with n-1 = 9 degrees of freedom and α = 0.05 (k = 3.666 and k = 10, for MDLs and MQLs, respectively), σb is the standard deviation of the procedural blank and m is the slope of the calibration curve. Intra-day and inter-day precisions were evaluated as RSD% of the peak area of each compound at two different levels, corresponding to the MQL (level I) and in the middle of the linearity range (level II). Intra-day and inter-day precisions were evaluated by three injections of level I and level II standard solutions within the same day and in three successive days, respectively. Linearity ranges were evaluated with at least seven concentration levels starting from MQL. Moreover, for real sample analysis, matrix effect (ME) was calculated according to the following equation [27].

$$ME= \left(\frac{m_{matrix}}{m_{solvent}} \times 100 \right) \text{-}100 \qquad (2)$$

where m_{matrix} and $m_{solvent}$ are the slopes of the calibration curves in matrix and 2 mM EDTA solution in MilliQ water, respectively.

4.3   Results and discussions

4.3.1   *Development and optimization of the multiresidue LC-MS/MS method*

The AQbD method development for the multiresidue, simultaneous LC–MS/MS analysis of the model group of PhCs and TPs followed a systematic workflow [30], consisting in the following AQbD phases: (i) analytical target profile (ATP) definition and method scouting; (ii) definition of CMAs, quality risk assessment and identification of CMPs; (iii) investigation of knowledge space by screening experimental design; (iv) RSM and definition of MODR.

4.3.1.1   *Analytical target profile and method scouting*

In this study, ATP, a set of objectives that the analytical method must achieve [32], consisted in developing an analytical method for the simultaneous analysis of a wide variety of analytes that allows obtaining the best compromise in terms of response factors, baseline resolutions, and analysis time. Within method scouting, a preliminary comparison between the C18 and PFP stationary phases was performed. In this regard, it should be underlined that the octadecyl silica phase has been widely used for the LC separation of PhCs [13, 33], whereas the pentafluorophenyl one has been much less applied to this issue, particularly for the development of methods dedicated to specific PhCs classes [9, 34]. The two stationary phases selected for this study are characterized by very different functionalization of silica particles, which translate into a wide and interesting range of interactions between target analytes and stationary phases themselves. More in detail, C18 stationary phase is mainly characterized by hydrophobic interactions, while PFP provides a much wider set of interactions, including π-π, hydrogen bonding, dipole-dipole and steric ones [35]. Different elution conditions were also tested in terms of ACN/MeOH ratio in the eluent "B", and acidity in the

aqueous mobile phase (HCOOH percentages 0% and 0.02%), at a flow rate of 0.45 mL/min, already found suitable for PhCs separation on the aforementioned stationary phases [9]. As a general outcome of the scouting, both stationary phases exhibited a satisfactory chromatographic behaviour for compounds detected under NI, even though the PFP showed a higher retention and chromatographic efficiency, ascribable to the presence of π-π interactions between aromatic moieties of the stationary phase and the analyte. Furthermore, the PFP column gave rise to a better resolution of the isobaric pairs 2-HYIBU/3-HYIBU and KET/FEN (see **Figure S2** in Section 4 of the *Supplementary material*), highlighting its remarkable positional isomer and shape selectivity related to the multiple interactions provided by this stationary phase. However, a significant improvement in the resolution of the above-mentioned critical isobaric pairs was observed on C18 by increasing the MeOH percentage in the organic eluent, due to its lower eluent strength towards these analytes, as evidenced by the increase in retention as the alcohol percentage increases. The addition of MeOH in the organic phase was also crucial for enhancing the ionization of estrogens, which however remained the compounds exhibiting the lowest sensitivity (see **Table S1**). This behaviour has been previously noticed [10] and dedicated methods have been recently proposed to meet the strict limits of European monitoring programs [36, 37]. Acidity was also a crucial parameter for estrogen sensitivity, since a decrease of the response factor of about one order of magnitude was generally observed by increasing HCOOH from 0% to 0.02%. Conversely, the presence of HCOOH was important for making narrower the peak of acidic analytes. Analytes detected under PI mode generally gave rise to broader peaks. This behaviour was particularly evident for the basic compounds AMX, ATE, and RNT when the elution was performed with ACN and 0.02% HCOOH in the aqueous mobile phase. In fact, under these experimental conditions, retention factors ≤0.5 and 1.1 were achieved with C18 and PFP, respectively. The use of ACN/MeOH mixtures produced a remarkable increase of retention factors on PFP column (1.3-2.5 depending on the analyte considered), whilst no significant change were observed

on C18. The divergent chromatographic behaviour and the very different sensitivity of targeted analytes as a function of the investigated parameters clearly highlighted the challenging character of this analytical study. Based on these findings, both stationary phases were included in the screening experiments. In addition, several parameters influencing chromatographic separations and response factors of PhCs were studied, as described below.

4.3.1.2 *Critical method attributes and critical method parameters*

To ensure and control the quality of the chromatographic method [38] the following CMAs were selected: (i) analysis time, (ii) chromatographic resolution, and (iii) response factor of the compounds reported in the "Watch List" 2018/840/EU [20]. Analysis time (t) was evaluated as the retention time of the last eluting peak. As regards chromatographic resolutions, during the screening phase, the separation of analyte pairs characterized by isobaric transitions, i.e. 2-HYIBU/3-HYIBU (R1) and KET/FEN (R2) were considered critical. Moreover, for MODR definition (i.e. RSM), five further critical resolutions were considered as potential border resolutions for polarity switching, to optimize the simultaneous acquisition of compounds detected under NI and those detected under PI. These resolutions were AZI/E3 (R3), E3/TAM (R4), TAM/2-HYIBU (R5), FUR/ERY (R6), and RAM/β-E2 (R7). Sensitivities were evaluated for EE2, β-E2, E1, CLA, AMX, AZI, ERY and CIP through S/N of the respective chromatographic peak. An Ishikawa diagram (see **Figure S3** of the Supplementary material) was drawn, aiming at pointing out the risk factors of the chromatographic method, thus identifying the CMPs which could influence the selected CMAs and should be in depth investigated by experimental design methodologies [39, 40]. Based on the analysis of the diagram and the results of the scouting, the following seven CMPs were considered: stationary phase (PHASE), ACN/MeOH ratio (RATIO), formic acid percentage (v/v) in the eluent A (HCOOH), initial organic phase percentage (v/v) (INOP), extent of the increase in the organic phase percentage (%B min^{-1}, INCOP), flow rate (mL min^{-1}, FLOW) and temperature (°C, T). The effects of the CMPs on the CMAs were evaluated by screening and RSM experimental designs.

4.3.1.3   Screening phase

Asymmetric/symmetric CMPs screening designs with 2–4 levels are generally employed, allowing a higher number of CMPs to be investigated with a lower number of experiments [41]. In this phase, the column stationary phase was investigated at two levels (i.e. C18 and PFP), while all the other six CMPs were studied at three levels (see **Table S4** of the *Supplementary material*). A Free-Wilson model [41], containing a constant term A0 and a number of coefficients for each factor corresponding to the number of levels of the factor minus one, was chosen (Equation 3):

$$y=A0+A1A+B1B+B2B+C1C+C2C+D1D+D2D+E1E+E2E+F1F+F2F+G1G+G2G \tag{3}$$

where: "A" is PHASE, "B" is RATIO, "C" is HCOOH%, "D" is INOP, "E" is INCOP, "F" is FLOW and "G" is T. As illustrated in **Table S5** of the *Supplementary material*, with this model, an asymmetric screening matrix consisting of 16 experiments (i.e. 2136//16) was obtained. Table S5 also shows the results obtained for the 11 CMAs in each experiment. The multiple linear regression (MLR) model defined by the investigated CMPs, correctly fitted the experimental data ($P < 0.05$) for all CMAs, except the S/N of EE2 and E1. For the latter variables, in fact, a significant decreasing trend of the response factor was observed only as a function of the increase of acidity of the eluent "A" (see **Figure S4** of the *Supplementary material*). Based on MLR results, R1 (see **Figure 1**), and S/N of CLA, AMX, and CIP (see **Figure 2**) were significantly improved by PFP stationary phase. Moreover, FLOW exerted a statistically significant effect on t, R1, R2 (**Fig. 1**), and S/N of CLA, AZI, ERY, and CIP (**Fig. 2**). More in detail, the experiments ran at 0.25 mL min^{-1} showed low selectivity, due to a general co-elution of all the analytes in the early stage of the chromatography. In addition, this flow rate translated in much lower sensitivities for macrolides and fluoroquinolones, and in longer analysis time (**Fig. 1-2**). Hence, PFP and 0.50 mL min^{-1} were fixed as PHASE and FLOW, respectively. Moreover, the experimental domain of the other factors was moved towards the zone leading to the best
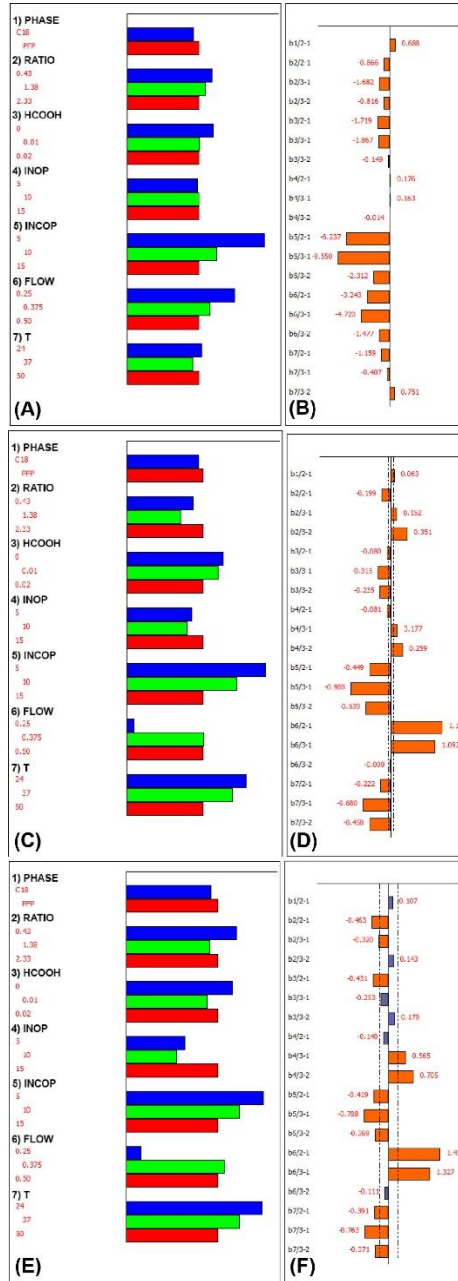
Figure 1 – Screening graphical analysis (A, C, E) and corresponding statistical significance (B, D, F) of critical method parameters (CMPs) on t (A, B), R1 (C, D), and R2 (E, F). The numbers associated to each CMP in boxes A, C, and E are related to the "b" indexes in boxes B, D, and F. For the meaning of CMP and CMA abbreviations, see paragraph 3.1.2 of the main text.

compromise for the optimization of the different CMAs. The change of the composition of the eluent "B" (i.e. RATIO) significantly affected most responses, even though with different trends. For example, R2 was maximized when MeOH was the predominant solvent of the organic phase (RATIO 0.43, **Fig. 1**). On the other hand, the sensitivities of macrolides exhibited an erratic trend with RATIO, since CLA and ERY showed increasing S/N values rising ACN percentage, whilst the increase of ACN in the eluent "B" from 58% (RATIO 1.38) to 70% (RATIO 2.33) gave rise to a statistically significant decrease in the sensitivity of AZI (**Fig. 2**). However, macrolides were among the most responsive PhCs, without any particular sensitive issue (**Table S1**). Due to this outlined complexity, RATIO was further investigated in the quite wide range 0.43-1.87. Concerning the percentage of HCOOH in the eluent "A", a very low or null acidity generally maximised the resolution of isobaric compounds (**Fig. 1**) and sensitivity of estrogens and AMX (see **Fig. 2** and **Fig. S4**), while macrolides and CIP were suppressed under these experimental conditions (**Fig. 2**). Due to this complex output, this CMP was further investigated by moving its domain to the range 0.005%-0.02%. The INOP affected significantly R1 and R2, being 15% the initial percentage of organic phase maximising the critical resolutions (**Fig. 1**). INOP influenced significantly also the S/N of AZI, ERY, and CIP with the highest sensitivity observed in between 10% and 15% of initial organic solvent (see **Fig. 2**). Accordingly, INOP was investigated in the successive RSM step within the narrower range of 10-15%. A more complex picture was observed for INCOP, which produced erratic responses depending on the CMA considered. More in detail, R1 and R2 (**Fig. 1**), as well as the sensitivity of AZI and AMX (**Fig. 2**) were maximised by medium-low INCOP values. However, for the further optimization step, a medium-high value (10-15%) of INCOP was selected to maximise the response factor of β-E2 (**Fig. 2**), which is one of the less ionisable target analyte, for which a high sensitivity is required by the European "Watch List" [20].
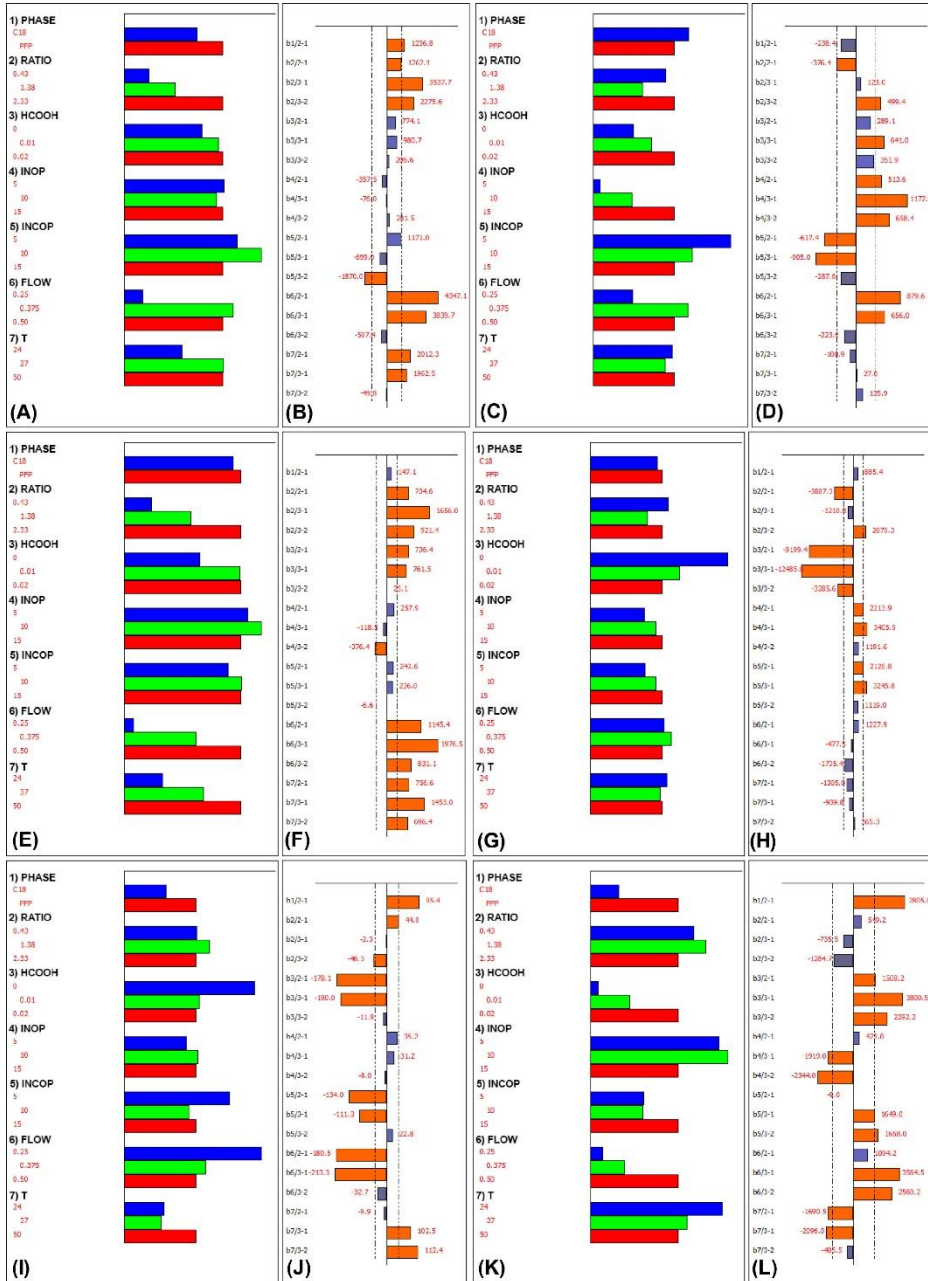
Figure 2 – Screening graphical analysis (A, C, E, G, I, K) and corresponding statistical significance (B, D, F, H, J, L) of critical method parameters (CMPs) on S/N of CLA (A, B), AZI (C, D), ERY (E, F), β-E2 (G, H), AMX (I, J), and CIP (K, L). The numbers associated to each CMP in boxes A, C, E, G, I, and K are related to the "b" indexes in boxes B, D, F, H, J, and L. For the meaning of CMP and CMA abbreviations, see paragraph 3.1.2 of the main text.

Finally, as regards T, medium-low values led to a significant improvement of R1, R2 (**Fig. 1**), and S/N of CIP, whereas for CLA, ERY, and AMX the highest sensitivities were achieved with medium-high temperatures (**Fig. 2**). Accordingly, as a good compromise, the new experimental domain of temperatures was moved towards medium-low values (i.e. 24-36 ℃).

4.3.1.4   *Response surface methodology and method operable design region*

RSM was carried out for obtaining predictive maps of the CMAs throughout the new selected experimental domain, with the ultimate goal of identifying MODR. It should be noted that the number of experiments increases compared to the screening phase, because at least three levels must be investigated for each CMP for verifying the presence of a curvature in the model [41]. Hence, to draw the maps, a polynomial quadratic model including linear, interaction and quadratic effects was postulated based on the five selected CMPs (**Table S4**), as reported in the Equation 4.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_{11}x_1^2+\beta_{22}x_2^2+\beta_{33}x_3^2+\beta_{44}x_4^2+\beta_{55}x_5^2+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{14}x_1x_4+\beta_{15}x_1x_5+\beta_{23}x_2x_3+\beta_{24}x_2x_4+\beta_{25}x_2x_5+\beta_{34}x_3x_4+\beta_{35}x_3x_5+\beta_{45}x_4x_5+\varepsilon \qquad (4)$$

where y is the experimental response, x_i are the independent variables, β_0 is the intercept, β_i are the true coefficients and ε is the experimental error. The coefficients of the model were calculated by using a Central Composite Orthogonal Design (CCO), composed of a Fractional Factorial Design and star points [42]. The experimental plan, reported in **Table S6**, was made by 29 experiments, including three replicates at the centre of the experimental domain. The total number of experiments was made up by three contributions: (i) 2k-1 (k = 5) fractional factorial experiments, (ii) 2×k (k = 5) axial experiments carried out on the axes at a distance ± α (1.6643) from the centre, calculated to obtain an orthogonal design, and (iii) replicates at the centre of the domain (n = 3) for the estimate of the experimental variance. In the development of this phase, starting from the previous set of CMAs, five additional critical resolutions were selected (see section 3.1.2). This further choice aimed at identifying potential pairs of

analytes, the resolutions of which may be used as an available scheduled window to set the polarity switch (following indicated as border resolutions), so as to achieve the simultaneous analysis of the 34 PhCs in a single chromatographic run. None of the selected responses were transformed and all the multiple linear regression (MLR) models were refined by deleting selected non-significant terms in order to improve the values of goodness of prediction (Q2). As illustrated in the **Table S7** of the *Supplementary material*, MLR statistics, performed with the ANOVA test, evidenced significant regression and validation models for all CMAs apart from t (significant only in regression) and S/N of EE2, ERY and CIP (not significant at all). Therefore, S/N of EE2, ERY and CIP were not included in RSM, whilst t was kept in RSM evaluation due to its very low experimental variance. The graphical analysis of the MLR coefficients reported in **Figure 3** allows highlighting the role of the linear, quadratic and interaction terms of each CMP, attempting to identify a region of the experimental domain that satisfies the optimal requirements of each CMA. RATIO and INCOP strongly affected the response t (**Fig. 3A**), evidencing that, moving to medium-high ranges, the analysis time can be reduced below the selected CMA threshold (i.e. 10 min). Moreover, the influence of the quadratic effects of the two factors evidenced a response-dip towards the upper boundaries of the investigated ranges (see **Figure S5** of the *Supplementary material*). It should be however noted that the experimental variability of analysis time was quite low, accounting it for about 1.5 min (**Table S6**). R1 was positively influenced by the increase of RATIO and INOP, whereas a negative effect was observed for INCOP and T (**Fig. 3B**). An opposite effect of INOP and INCOP was observed also for R2, which was negatively affected by the INOP quadratic term and its interaction with RATIO (**Fig. 3C**). RATIO, HCOOH, and their quadratic coefficients greatly influenced most border resolutions, even though with different effects on the separations of specific pairs (**Fig. 3D-H**). Indeed, **Fig. 3** shows very different dependencies of the seven critical resolutions by CMPs, highlighting the challenging character of satisfying even only some of the resolutions identified as critical. Eluent acidity was also

confirmed as a significant suppressive factor for the response of β-E2 and E1, the former being negatively influenced also by T (**Fig. 3I-J**). Conversely, the temperature increase enhanced the signal of CLA, which was also increased by increasing methanol in the eluent "B" (**Fig. 3K**). The sensitivity of AMX took advantage of low values of INOP and INCOP (**Fig. 3L**), whereas for S/N of AZI, temperature exerted a negative effect and a multiple positive dependence was observed for INOP, INCOP, and HCOOH (**Fig. 3M**). Such a complex scenario does not permit to identify an optimal CMP experimental region satisfying all the investigated CMAs (i.e. the sweet spot). Based on the results reported in **Table S6**, among the 13 investigated CMAs, t, R3, R4, and S/N of CLA and AZI were excluded from the sweet spot modelling, owing to their low variability (i.e. t and S/N of CLA and AZI) or a high risk of failure (i.e. R3 and R4). It should also be underlined that the inclusion of R1 and R2 is mandatory due to their isobaric behaviour, and a minimum target value of 1.00 was adopted, since it usually corresponds to the pseudo-baseline resolution (i.e. 2% of peak overlapping). Similarly, for R5-R7 resolutions a minimum threshold limit of 1.00 was selected. Moreover, the selection of the aforementioned S/N response variables is relevant, due to their inclusion in the 2018/840/EU decision and the low response factors observed in comparison with the other PhCs (**Table S1**). However, the identification of proper values for the minimum S/N requirements of these analytes is not easy at all, since E1 and β-E2 have a maximum acceptable detection limit in the 2018/840/EU decision of 0.4 ng L^{-1}, i.e. about two orders of magnitude lower than AMX (78 ng L^{-1}). In this regard, it should be emphasized that, among the three estrogens investigated, EE2 has a sensitivity even lower than that of E1 and β -E2 (see **Table S1**) and a maximum acceptable detection limit an order of magnitude lower than the other two estrogens. On the other hand, the highest S/N values determined for AMX under the RSM experimental conditions were approximately 30-50 times lower than the corresponding estrogens. Hence, it is considered appropriate to set the required targets for the S/N of AMX and estrogens equal to 150 and 1500 respectively. The 4D sweet spot plots obtained

considering the aforementioned CMA threshold limits are reported in **Figure S6** of the *Supplementary material*, as the light blue areas where the desired values for all the eight CMAs were reached. Three sweet spots were identified at the INOP vs. INCOP values of 12.5-12.5, 15-12.5, and 12.5-10, covering an HCOOH percentage of 0.01-0.02 and a RATIO interval of about 0.4-1.0. The MODR was then calculated by a Monte Carlo simulations for risk analysis [31], by setting the probability level for the simultaneous fulfil of all targets as $\geq 90\%$, thus obtaining the risk map reported in **Figure 4**. The obtained MODR highlighted the presence of various experimental domains characterized by probability of failure $\leq 5\%$. In this regard, it should be noted that these domains include but were not limited to the sweet spots reported in **Fig. S6**, due to the uncertainties associated to the CMAs used for MODR definition. Five RATIO and HCOOH conditions (i.e. 0.75-0.010%, 0.79-0.008%, 1.69-0.009%, 1.75-0.008%, and 1.76-0.011%, see **Fig. 4**) included in the 5% failure probability regions and comprising the sweet spots, were tested at T=30 °C, INOP=12.5 %, and INCOP of 12.5 %B/min. Among these experimental conditions, the RATIO-HCOOH combination 1.76-0.011% satisfied the mandatory resolutions R1 and R2 and gave rise to the best compromise in terms of AMX and estrogen sensitivities, as well as of number of scheduled CPS acquisition windows (see **Figure 5**). Under these optimized conditions, among the five resolutions included in the MODR definition, only E3/TAM and RAM/β-E2 were achieved. However, the optimization process allowed for scheduling three acquisition windows with single polarity and four with CPS (see **Table S3** for full details). Hence, the CPS was applied only to a few transitions for each polarity switching window, thus achieving signal acquisitions much more accurate than those obtained using the CPS over the entire chromatographic run, and approximately identical to the single polarity acquisition mode. **Figure S7** of the *Supplementary material* illustrates this behaviour for DZP and DIC, as representative examples of basic and acidic analytes detectable under PI and NI modes, respectively.
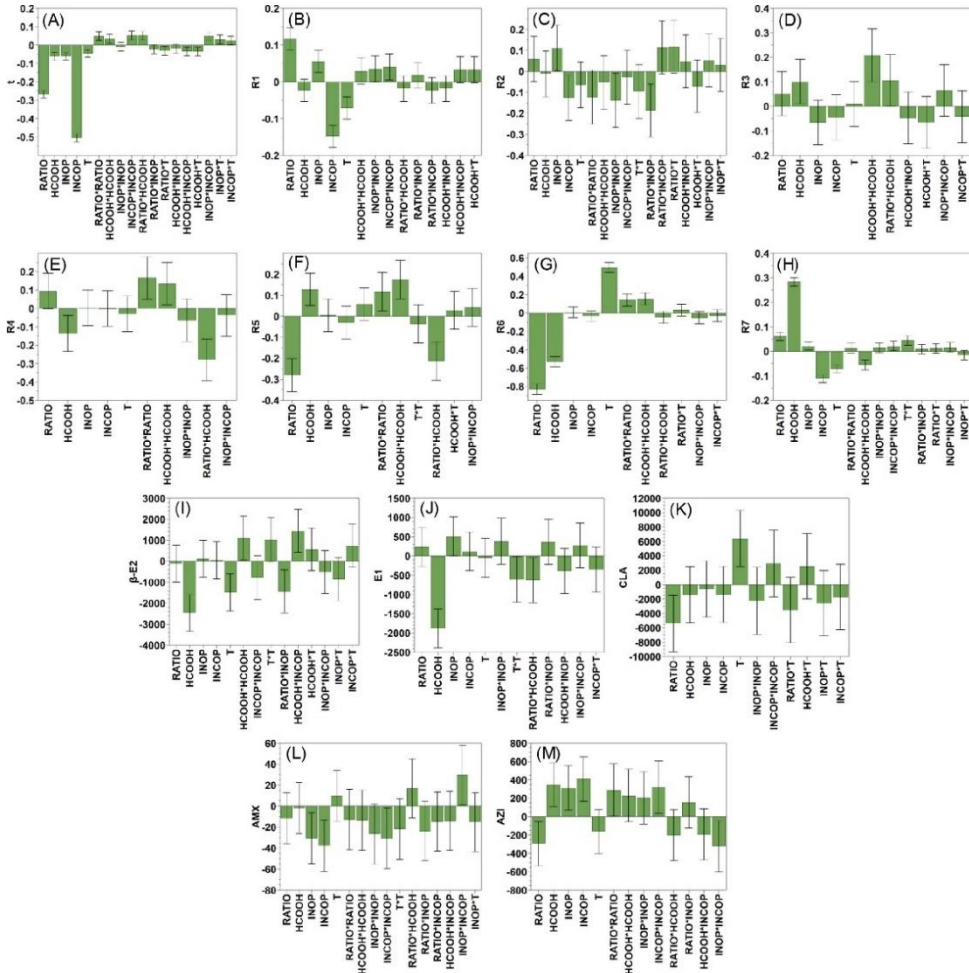
Figure 3 – Graphical analysis of the coefficients of the critical method parameters (CMPs) used for obtaining the multiple linear regression models of each investigated critical method attributes (CMAs) within the response surface methodology phase. For the meaning of CMP and CMA abbreviations, see paragraph 3.1.2 of the main text.
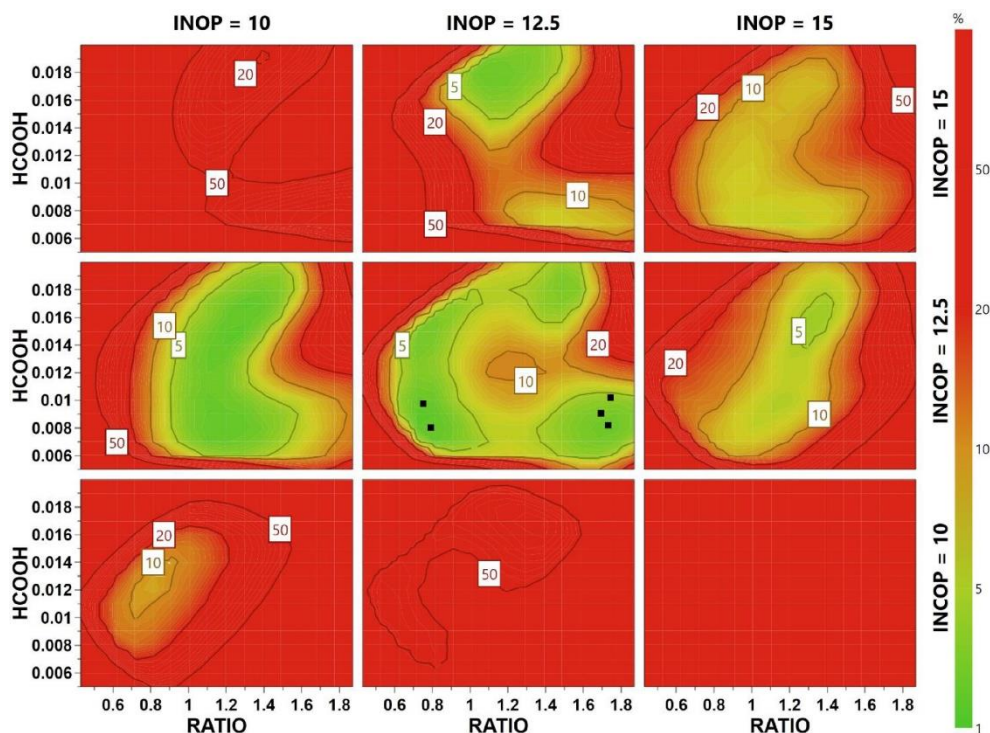
Figure 4 – Risk maps obtained by plotting RATIO vs. HCOOH at three different levels of INOP (10%, 12.5%, and 15%) and INCOP (10%, 12.5%, and 15% B/min). The thresholds for the risk of failure of 10% are the following: R1, R2, R5, R6, and R7 resolution values <1, S/N of AMX <150 and S/N of β-E2 and E1 <1500. Temperature was set at 30°C. For the meaning of CMP and CMA abbreviations, see paragraph 3.1.2 of the main text. Squared symbols represent the experimental points tested.

4.3.2   *Figures of merit of the LVDI method*

The performances of the optimized method are illustrated in **Table 1**, whereas the total ion current chromatogram is reported in **Figure S8** of the Supplementary material. The method showed good precision as measured by RSD %, being intra-day values in the range of 0.2-8.1% and 0.6-8.9% at levels I and II, respectively. Comparable or slightly higher RSD % were observed for inter-day precision. Calibration lines were in all cases characterized by high determination coefficients ($R^2 \geq 0.9902$) and linearity ranges covering at least two orders of magnitude, with the
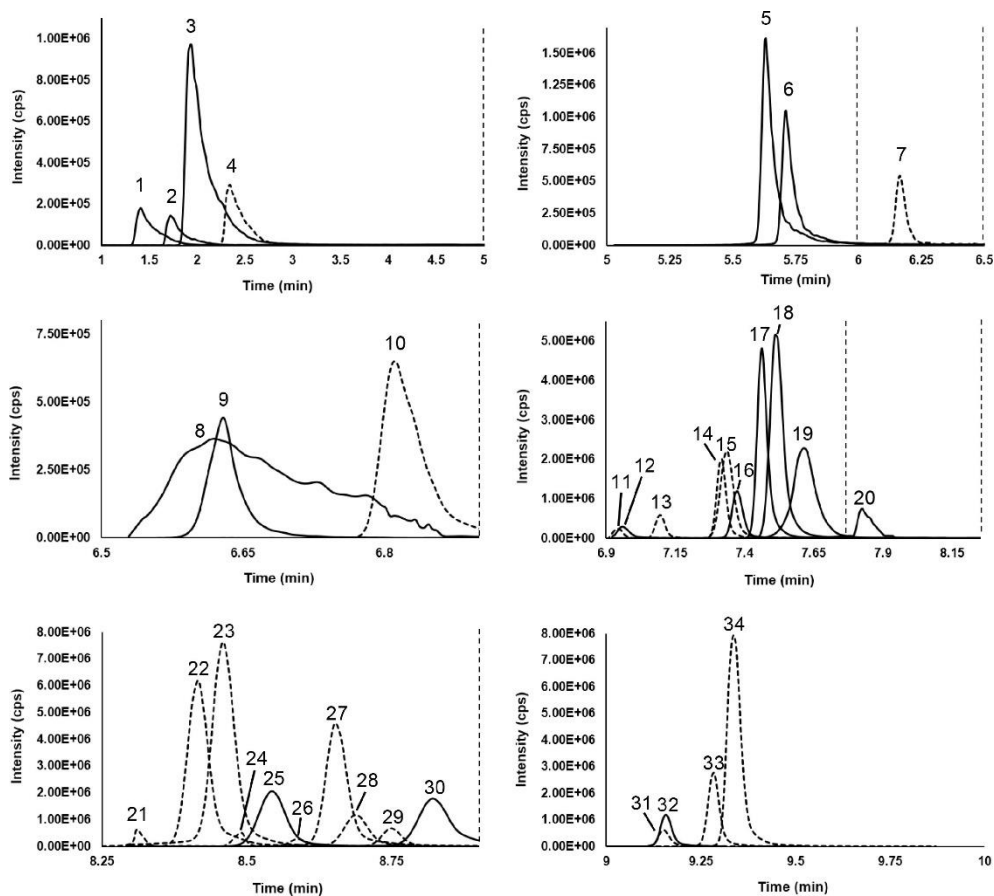
Figure 5 – MRM acquisition windows of the investigated group of pharmaceutical compounds and transformation products in the optimized experimental conditions. Solid and dashed chromatographic traces refer to acquisitions under positive and negative ionization, respectively. (1) AMX; (2) ATE; (3) RNT; (4) PAR; (5) LVF; (6) CIP; (7) ASA; (8) AZI; (9) BIS; (10) E3; (11) 2-HYIBU; (12) TAM; (13) 3-HYIBU; (14) 1-HYIBU; (15) O-DMANP; (16) CBZ; (17) PAN; (18) FUR; (19) ERY; (20) RAM; (21) α-E2; (22) KET; (23) 4'-HYDIC; (24) β-E2; (25) DZP; (26) EE2; (27) FEN; (28) NAP; (29) E1; (30) CLA; (31) ATO; (32) FLU; (33) IBU; (34) DIC. The meaning of compound abbreviations is reported in Table S1.

only exceptions of CIP, 2-HYBU, and 3-HYBU, whose response was linear over one order of magnitude. Interestingly, MDLs of macrolides (1-16 ng L^{-1}) were in line with the detection limits requested by the "Watch List" 2018/840/EU (19 ng L^{-1}). Very high sensitivities, from sub-ng L^{-1} to low ng L^{-1}, were found for most target analytes not included in the aforementioned "Watch List" (i.e. RNT, ATE,

ASA, BIS, AZI, CBZ, TAM, PAN, 2-HYIBU, 3-HYIBU, 1-HYIBU, O-DMNAP, FUR, DZP, NAP, ATO, and FLU), whilst nine analytes (AMX, ACE, LVF, RMP, 4'-HYDIC, KET, FEN, DIC, and IBU) showed MDLs of tens-hundreds of ng L^{-1}. MDLs in the low µg L^{-1} range were found for CIP and estrogens apart from E3 and EE2, which exhibited sensitivities of tens of µg L^{-1}. This latter finding clearly highlights the challenging character of the quantification of estrogens, the determination of which requires dedicated off-line or on-line solid phase extraction (SPE) methods when sensitivities approaching or satisfying the strict detection limits reported in the European "Watch List" must be achieved [10, 36, 43]. The sensitivities obtained in this study can be compared with those reported in the few researches adopting LVDI for PhCs analysis (**Table S8**) [8, 44, 45]. Based on this comparison, limits here obtained resulted better than or comparable to those achieved elsewhere, with the main exceptions of some acidic compounds (i.e. KET, FEN, NAP, DIC, and IBU), the sensitivities of which were probably penalized during the multi-compound AQbD optimization, which pursues the search for a compromise.

Table 1 – Method detection (MDLs) and quantitation (MQLs) limits (µg L^{-1}), intra-day (RSDintra, %) and inter-day (RSDinter, %) precision of peak areas at concentration levels corresponding to S/N values of IQLs (Level I) and to central points of linearity ranges (Level II) obtained by injecting standard solutions in MilliQ water. Linearity data: slope of calibration curves (m), determination coefficients (R2), and linearity ranges (µg L^{-1}).

Compound	MDL	MQL	Level I		Level II		S/N	m	R^2	Linearity range
			RSD$_{intra}$	RSD$_{inter}$	RSD$_{intra}$	RSD$_{inter}$				
RNT	0.0003	0.0007	2.9	6.7	3.1	7.5	10-200	624790	0.9997	0.0007-10
ATE	0.002	0.006	3.3	5.4	0.7	3.4	10-200	169051	0.9987	0.006-5
AMX	0.4	1.0	0.5	10.4	1.3	5.3	10-200	908	0.9991	1.0-250
ACE	0.01	0.03	0.9	1.2	3.0	6.9	10-200	12927	0.9992	0.03-50
LVF	0.04	0.10	3.2	10.6	5.3	10.0	10-50	174537	0.9975	0.10-12.5
CIP	1	3	0.2	0.7	2.5	6.4	10-200	21475	0.9930	3-25
ASA	0.03	0.007	1.6	4.1	3.4	10.8	10-200	49183	0.9982	0.007-0.5
BIS	0.0002	0.0007	2.6	10.5	4.6	5.1	10-200	1966133	0.9990	0.0007-0.25
AZI	0.006	0.02	2.7	5.4	1.2	6.8	10-200	12004	0.9926	0.02-2
E3	25	68	2.8	7.9	3.6	6.1	10-50	90	0.9956	68-5000
CBZ	0.0001	0.0003	0.9	3.9	2.2	4.2	25-200	1648149	0.9903	0.0003-1
RMP	0.01	0.03	1.8	4.3	1.7	10.2	10-200	150309	0.9999	0.03-7.5
TAM	0.001	0.002	0.6	2.5	0.9	3.2	10-200	1290808	0.9906	0.002-0.25
PAN	0.01	0.02	0.4	10.8	1.5	10.3	10-200	442211	0.9951	0.02-4
ERY	0.02	0.04	2.6	7.9	3.6	10.4	10-200	608095	0.9992	0.04-2.5
2-HYIBU	0.02	0.05	0.7	2.4	4.9	5.2	10-200	339354	0.9992	0.05-0.5
3-HYIBU	0.02	0.05	5.0	10.8	3.0	8.5	10-200	85274	0.9997	0.05-1.5
1-HYIBU	0.02	0.05	8.1	9.6	5.4	9.4	10-200	50392	0.9993	0.05-8
O-DMNAP	0.01	0.04	4.3	6.4	4.0	6.2	10-200	108864	0.9966	0.04-8
FUR	0.01	0.02	1.7	3.1	2.6	4.4	10-50	204521	0.9939	0.02-25
CLA	0.001	0.003	2.8	10.1	1.9	7.7	10-200	489530	0.9975	0.003-0.5
DZP	0.01	0.03	1.3	10.8	3.1	4.5	10-200	395076	0.9975	0.03-1
4'-HYDIC	0.1	0.3	0.2	3.0	2.9	5.3	10-200	168858	0.9905	0.3-15
KET	0.1	0.2	1.1	10.3	4.4	9.0	10-200	117494	0.9925	0.2-15
FEN	0.08	0.21	1.8	3.2	2.9	5.8	10-200	143011	0.9965	0.2-8
NAP	0.01	0.03	1.6	5.7	5.1	8.5	10-200	49376	0.9980	0.03-8
EE2	32	95	3.5	10.1	5.1	10.4	50-200	86	0.9903	95-4000
α-E2	4	10	2.4	5.4	5.6	10.0	50-200	27	0.9902	10-1750
β-E2	2	7	1.4	8.5	2.7	6.9	50-200	73	0.9945	7-1750
E1	1	4	0.8	9.0	1.5	9.5	10-200	433	0.9936	4-750
ATO	0.0002	0.0005	0.6	1.2	1.8	6.7	10-200	515808	0.9996	0.0005-0.5
DIC	0.04	0.1	2.4	3.1	2.3	7.7	10-200	188374	0.9903	0.1-12.5
IBU	0.09	0.3	2.4	10.7	8.9	10.4	10-200	49291	0.9986	0.3-15
FLU	0.01	0.02	3.2	5.0	0.6	3.4	10-200	70756	0.9997	0.02-2

4.3.3   Application to the analysis of real samples

The optimized method was applied to the analysis of PhCs and TPs in the effluents of four WWTPs and in SW samples collected in the close proximity of their outlets. ME%, calculated for each target analyte in each investigated sample according to the Equation 2, are reported in **Figure S9** of the *Supplementary material*. The influence of the matrix on the mass spectrometric signal was significant, being │ME%│far higher than 20% in almost all cases. Hence, the matrix matched calibration or the standard addition method must be adopted for the accurate determination of target analytes in real samples. In this regard, it should be noted that for most target analytes and investigated samples the signal was suppressed, thus lowering the sensitivity. On the contrary, in all the investigated samples, a signal amplification was observed for the fluoroquinolones LVF and CIP, as well as for the AZI, the latter showing therefore an opposite behaviour with respect to ERY and CLA. As illustrated in **Table 2**, seventeen PhCs and two TPs were quantified in at least one of the investigated samples, in a wide range of concentrations, approximately from 1 to 3200 ng L^{-1}. As a whole, effluents from municipal WWTPs (i.e. TWW3 and TWW4), which receives also hospital wastewater, showed PhCs concentrations about two times higher than those determined in the outlet of the mixed industrial-municipal facilities (TWW1 and TWW2), highlighting the role of domestic sewages in the spreading of contamination by PhCs. Among target analytes, LVF, FUR, KET, and IBU were found in both TWWs and SWs at higher concentration values than the others, ranging between hundreds of ng L^{-1} and few µg L^{-1}. Indeed, these analytes were among the most consumed drugs by the Italian population [19]. In detail, LVF occurred the most in TWW3 (1060 ng L^{-1}) and TWW4 (1239 ng L^{-1}) samples. An analogous result was evidenced for FUR, which was found in TWW3 and TWW4 at concentration values of 1647 and 3234 ng L^{-1}, respectively, but also in TWW1 (1016 ng L^{-1}), probably due to the larger domestic contribution to the mixed industrial-municipal wastewater in WWTP1 compared to WWTP2 (see **Table S2**). As regards non-steroidal anti-inflammatory drugs (NSAIDs), IBU and

KET occurred in TWWs at concentration values (886-3232 ng L^{-1} and 295-1104 ng L^{-1}) much higher than DIC (<50-936 ng L^{-1}), notwithstanding the latter is reported as the most prescribed/consumed NSAID in Italy [19] and poorly degradable compared to IBU [46]. Interestingly, the same relative abundance of these NSAIDs was previously reported in the sludge of these WWTPs [9]. Hence, the concentrations found for these NSAIDs probably reflected their actual consume in the investigated area. Among the analytes included in the European "Watch List" 2018/840/EU, only two antibiotics were determined. More in detail, AZI occurred in all the investigated TWWs and SWs at the concentration ranges of 84-350 ng L^{-1} and 40-59 ng L^{-1}, respectively, whereas ERY was detected in all TWWs, but quantified only in TWW4 (72 ng L^{-1}). These findings are in agreement with the overall scenario described by the consumption data of PhCs in Italy, which evidenced that macrolides in general and especially AZI are the most common prescribed antibiotics in hospital setting, particularly in paediatric patients [19].

4.4   Conclusions

The AQbD optimization strategy here adopted for the analysis of a wide group of chemically heterogeneous PhCs and TPs proved to be essential in managing the high number of critical method factors together with the several dependent variables investigated (i.e. analysis time, nine resolutions and eighteen S/N). In fact, such multivariate approach allowed to identify the region of the experimental domain where both chromatographic resolutions and sensitivities could be maximized. In this way, it was also possible to schedule single-polarity and polarity-switched acquisition windows dedicated to specific compounds, maximizing the overall method sensitivity. The application of the method to TWW and SW samples highlighted its throughput in terms of analysis time (i.e. <10 min per sample) and therefore its suitability for a rapid screening of the presence of PhCs and TPs in aqueous matrices, notwithstanding the quite high matrix effects. Furthermore, although the detection limits achieved here are not low enough for monitoring some PhCs and especially estrogens in accordance

with the current European "Watch List" by LVDI, the use of this optimized method after on-line or off-line SPE with adequate preconcentration factors, can overcome this problem. Even though this study was not designed as an environmental monitoring of target compounds and included only a few TWW and SW samples collected in a brief period, it is remarkable that 19 out the 34 target analytes were detected, revealing in some cases a quite good agreement between drug prescription/consumption and presence in the investigated samples. Furthermore, many PhCs and TPs were determined not only in TWWs, but also in SWs, notwithstanding the dilution effect provided by the receiving water bodies, thus evidencing the impact of WWTPs on the PhCs release in the surrounding environment, due to their high consumption and recalcitrance to purification processes.

Table 2 – Concentration values (ng/L) of the pharmaceutical compounds (PhCs) and transformation products found in the investigated treated wastewaters (TWWs) and surface waters (SWs). For full details of sampling points and sample characteristics, see Table S1 of the Supplementary material.

PhCs	TWW1	TWW2	TWW3	TWW4	SW1	SW2	SW3	SW4
RNT	<0.6 [a]	41	62	49	<0.4 [a]	<0.3 [a]	<0.3 [a]	<0.3 [a]
ATE	164	52	65	142	40	50	10	30
LVF	202	795	1060	1239	100	570	830	800
CIP	310-860 [b]	200-550 [b]	160-420 [b]	150-420 [b]	<250 [a]	450-1220 [b]	400-1080 [b]	190-510 [b]
BIS	34	7	149	217	0.007	8	3	15
AZI	350	84	171	312	50	40	59	50
CBZ	241	10	921	794	50	40	10	150
RMP	274	30	139	191	<10	10-20 [b]	10-30 [b]	60
TAM	3.3	<2 [a]	3.4	<2 [a]	<1 [a]	<1 [a]	3	<1 [a]
ERY	180-490 [b]	90-250 [b]	90-250 [b]	72	<90 [a]	<110	<90 [a]	<100 [a]
2-HYIBU	370	610	310	170	160	60	180	150
FUR	1016	865	1647	3234	720	230	500	880
4'-HYDIC	<430 [a]	<110 [a]	<150 [a]	316	<110 [a]	130-350	610	120-320 [b]
KET	650	522	1104	295	270	490	460	290
FEN	160-430 [b]	130-350 [b]	130-350 [b]	<80 [a]	<110 [a]	990-2270 [b]	<130 [a]	<120 [a]
NAP	<20 [a]	<20 [a]	20-60 [b]	10-30 [b]	<20 [a]	<20 [a]	<20 [a]	<20 [a]
ATO	4.3	<0.1 [a]	<0.1 [a]	11	<0.1 [a]	0.9	1	<0.1 [a]
DIC	<60 [a]	<50 [a]	148	936	50-140 [b]	220	30-150 [b]	50-150 [b]
IBU	886	1840	3232	3089	690	1440	1190	910

[a] Method detection limit in real samples.
[b] Calculated concentration value comprised between method and quantitation limits in real samples.

Section 4

References

[1] M. Farré, L. Kantiani, M. Petrovic, S. Pérez, D. Barceló, Achievements and future trends in the analysis of emerging organic contaminants in environmental samples by mass spectrometry and bioanalytical techniques, Journal of chromatography A, 1259 (2012) 86-99 https://doi.org/10.1016/j.chroma.2012.07.024.

[2] S. Castiglioni, E. Davoli, F. Riva, M. Palmiotto, P. Camporini, A. Manenti, E. Zuccato, Mass balance of emerging contaminants in the water cycle of a highly urbanized and industrialized area of Italy, Water research, 131 (2018) 287-298 https://doi.org/10.1016/j.watres.2017.12.047.

[3] A. Pal, K.Y.-H. Gin, A.Y.-C. Lin, M. Reinhard, Impacts of emerging organic contaminants on freshwater resources: review of recent occurrences, sources, fate and effects, Science of the total environment, 408 (2010) 6062-6069 https://doi.org/10.1016/j.scitotenv.2010.09.026.

[4] J. Wang, L. Chu, L. Wojnárovits, E. Takács, Occurrence and fate of antibiotics, antibiotic resistant genes (ARGs) and antibiotic resistant bacteria (ARB) in municipal wastewater treatment plant: An overview, Science of The Total Environment, (2020) 140997 https://doi.org/10.1016/j.scitotenv.2020.140997.

[5] E. Archer, B. Petrie, B. Kasprzyk-Hordern, G.M. Wolfaardt, The fate of pharmaceuticals and personal care products (PPCPs), endocrine disrupting contaminants (EDCs), metabolites and illicit drugs in a WWTW and environmental waters, Chemosphere, 174 (2017) 437-446 https://doi.org/10.1016/j.chemosphere.2017.01.101.

[6] A. Gogoi, P. Mazumder, V.K. Tyagi, G.T. Chaminda, A.K. An, M. Kumar, Occurrence and fate of emerging contaminants in water environment: a review, Groundwater for Sustainable Development, 6 (2018) 169-180 https://doi.org/10.1016/j.gsd.2017.12.009.

[7] P. Verlicchi, M. Al Aukidy, E. Zambello, Occurrence of pharmaceutical compounds in urban wastewater: removal, mass load and environmental risk after a secondary treatment—a review, Science of the total environment, 429 (2012) 123-155 https://doi.org/10.1016/j.scitotenv.2012.04.028.

[8] T.S. Oliveira, M. Murphy, N. Mendola, V. Wong, D. Carlson, L. Waring, Characterization of pharmaceuticals and personal care products in hospital effluent and waste water influent/effluent by direct-injection LC-MS-MS, Science of the total environment, 518 (2015) 459-478 https://doi.org/10.1016/j.scitotenv.2015.02.104.

[9] D. Rossini, L. Ciofi, C. Ancillotti, L. Checchini, M.C. Bruzzoniti, L. Rivoira, D. Fibbi, S. Orlandini, M. Del Bubba, Innovative combination of QuEChERS extraction with on-line solid-phase extract purification and pre-concentration, followed by liquid chromatography-tandem mass spectrometry for the determination of non-steroidal anti-inflammatory drugs and their metabolites in sewage sludge, Analytica chimica acta, 935 (2016) 269-281 https://doi.org/10.1016/j.aca.2016.06.023.

[10] L. Ciofi, D. Fibbi, U. Chiuminatto, E. Coppini, L. Checchini, M. Del Bubba, Fully-automated on-line solid phase extraction coupled to high-performance liquid chromatography–tandem mass spectrometric analysis at sub-ng/L levels of selected estrogens in surface water and wastewater, Journal of Chromatography A, 1283 (2013) 53-61 https://doi.org/10.1016/j.chroma.2013.01.084.

[11] M. Gros, M. Petrović, D. Barceló, Multi-residue analytical methods using LC-tandem MS for the determination of pharmaceuticals in environmental and wastewater samples: a review, Analytical and bioanalytical chemistry, 386 (2006) 941-952 https://doi.org/10.1007/s00216-006-0586-z.

[12] T.Y. Fang, S.M. Praveena, C. deBurbure, A.Z. Aris, S.N.S. Ismail, I. Rasdi, Analytical techniques for steroid estrogens in water samples-A review, Chemosphere, 165 (2016) 358-368 https://doi.org/10.1016/j.chemosphere.2016.09.051.

[13] M. Gros, S. Rodríguez-Mozaz, D. Barceló, Fast and comprehensive multi-residue analysis of a broad range of human and veterinary pharmaceuticals and some of their metabolites in surface and treated waters by ultra-high-performance liquid chromatography coupled to quadrupole-linear ion trap tandem mass spectrometry, Journal of Chromatography A, 1248 (2012) 104-121 https://doi.org/10.1016/j.chroma.2012.05.084.

[14] B. Petrie, J. Youdan, R. Barden, B. Kasprzyk-Hordern, Multi-residue analysis of 90 emerging contaminants in liquid and solid environmental matrices by ultra-high-performance liquid chromatography tandem mass spectrometry, Journal of Chromatography A, 1431 (2016) 64-78 https://doi.org/10.1016/j.chroma.2015.12.036.

[15] M. Zhong, T. Wang, C. Qi, G. Peng, M. Lu, J. Huang, L. Blaney, G. Yu, Automated online solid-phase extraction liquid chromatography tandem mass spectrometry investigation for simultaneous quantification of per-and polyfluoroalkyl substances, pharmaceuticals and personal care products, and organophosphorus flame retardants in environmental waters, Journal of Chromatography A, 1602 (2019) 350-358 https://doi.org/10.1016/j.chroma.2019.06.012.

[16] M.J. García-Galán, M. Petrovic, S. Rodríguez-Mozaz, D. Barceló, Multiresidue trace analysis of pharmaceuticals, their human metabolites and transformation products by fully automated on-line solid-phase extraction-liquid chromatography-tandem mass spectrometry, Talanta, 158 (2016) 330-341 https://doi.org/10.1016/j.talanta.2016.05.061.

[17] L. Ciofi, C. Ancillotti, U. Chiuminatto, D. Fibbi, L. Checchini, S. Orlandini, M. Del Bubba, Liquid chromatographic–tandem mass spectrometric method for the simultaneous determination of alkylphenols polyethoxylates, alkylphenoxy carboxylates and alkylphenols in wastewater and surface-water, Journal of Chromatography A, 1362 (2014) 75-88 https://doi.org/10.1016/j.jchromb.2008.10.018.

[18] N. Hermes, K.S. Jewell, A. Wick, T.A. Ternes, Quantification of more than 150 micropollutants including transformation products in aqueous samples by liquid chromatography-tandem mass spectrometry using scheduled multiple

reaction monitoring, Journal of Chromatography A, 1531 (2018) 64-73 https://doi.org/10.1016/j.chroma.2017.11.020.

[19] Italian Medicines Agency, The Medicines Utilisation Monitoring Centre. National Report on Medicines use in Italy: Year 2019. Available on line at https://www.aifa.gov.it/documents/20142/1205984/rapporto-osmed-2019.pdf/f41e53a4-710a-7f75-4257-404647d0fe1e, (2020).

[20] COMMISSION IMPLEMENTING DECISION (EU) 2018/840 of 5 June 2018 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council and repealing Commission Implementing Decision (EU) 2015/495, Official Journal of the European Union, 2018, pp. 9-12.

[21] B. Pasquini, S. Orlandini, S. Furlanetto, R. Gotti, M. Del Bubba, F. Boscaro, B. Bertaccini, M. Douša, G. Pieraccini, Quality by Design as a risk-based strategy in pharmaceutical analysis: Development of a Liquid chromatography-Tandem mass spectrometry method for the determination of Nintedanib and its impurities, Journal of Chromatography A, 1611 (2020) 460615 https://doi.org/10.1016/j.chroma.2019.460615.

[22] B. Pasquini, S. Orlandini, M. Villar-Navarro, C. Caprini, M. Del Bubba, M. Douša, A. Giuffrida, R. Gotti, S. Furlanetto, Chiral capillary zone electrophoresis in enantioseparation and analysis of cinacalcet impurities: Use of Quality by Design principles in method development, Journal of Chromatography A, 1568 (2018) 205-213 https://doi.org/10.1016/j.chroma.2018.07.021.

[23] L. Nompari, S. Orlandini, B. Pasquini, C. Campa, M. Rovini, M. Del Bubba, S. Furlanetto, Quality by design approach in the development of an ultra-high-performance liquid chromatography method for Bexsero meningococcal group B vaccine, Talanta, 178 (2018) 552-562 https://doi.org/10.1016/j.talanta.2017.09.077.

[24] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, M. Douša, S. Pinzauti, S. Furlanetto, Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: visualizing the design space within quality by design framework, Journal of Chromatography A, 1467 (2016) 363-371 https://doi.org/10.1016/j.chroma.2016.06.082.

[25] C. Ancillotti, S. Orlandini, L. Ciofi, B. Pasquini, C. Caprini, C. Droandi, S. Furlanetto, M. Del Bubba, Quality by design compliant strategy for the development of a liquid chromatography–tandem mass spectrometry method for the determination of selected polyphenols in Diospyros kaki, Journal of Chromatography A, 1569 (2018) 79-90 https://doi.org/10.1016/j.chroma.2018.07.046.

[26] P. Silva, C.L. Silva, R. Perestrelo, F.M. Nunes, J.S. Câmara, A useful strategy based on chromatographic data combined with quality-by-design approach for food analysis applications. The case study of furanic derivatives in sugarcane honey, Journal of Chromatography A, 1520 (2017) 117-126 https://doi.org/10.1016/j.chroma.2017.09.019.

[27] L. Ciofi, L. Renai, D. Rossini, C. Ancillotti, A. Falai, D. Fibbi, M.C. Bruzzoniti, J.J. Santana-Rodriguez, S. Orlandini, M. Del Bubba, Applicability of

the direct injection liquid chromatographic tandem mass spectrometric analytical approach to the sub-ng L−1 determination of perfluoro-alkyl acids in waste, surface, ground and drinking water samples, Talanta, 176 (2018) 412-421 https://doi.org/10.1016/j.talanta.2017.08.052.

[28] R. Mirzaei, M. Yunesian, S. Nasseri, M. Gholami, E. Jalilzadeh, S. Shoeibi, H.S. Bidshahi, A. Mesdaghinia, An optimized SPE-LC-MS/MS method for antibiotics residue analysis in ground, surface and treated water samples by response surface methodology-central composite design, Journal of Environmental Health Science and Engineering, 15 (2017) 21 https://doi.org/10.1186/s40201-017-0282-2.

[29] J. Cha, S. Yang, K. Carlson, Trace determination of β-lactam antibiotics in surface water and urban wastewater using liquid chromatography combined with electrospray tandem mass spectrometry, Journal of Chromatography A, 1115 (2006) 46-57 https://doi.org/10.1016/j.chroma.2006.02.086.

[30] S. Orlandini, S. Pinzauti, S. Furlanetto, Application of quality by design to the development of analytical separation methods, Analytical and bioanalytical chemistry, 405 (2013) 443-450 https://doi.org/10.1007/s00216-012-6302-2.

[31] M.A. Herrador, A.G. Asuero, A.G. González, Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: An overview, Chemometrics and intelligent laboratory systems, 79 (2005) 115-122 https://doi.org/10.1016/j.chemolab.2005.04.010.

[32] ICH Harmonised Tripartite Guideline, Pharmaceutical Development Q(R2), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, (2009).

[33] A. Jelic, M. Gros, A. Ginebreda, R. Cespedes-Sànchez, F. Ventura, M. Petrovic, D. Barceló, Occurrence, partition and removal of pharmaceuticals in sewage water and sludge during wastewater treatment, Water research, 45 (2011) 1165-1176 https://doi.org/10.1016/j.watres.2010.11.010.

[34] S. Needham, P. Brown, K. Duff, D. Bell, Optimized stationary phases for the high-performance liquid chromatography–electrospray ionization mass spectrometric analysis of basic pharmaceuticals, Journal of Chromatography A, 869 (2000) 159-170 https://doi.org/10.1016/j.watres.2010.11.010.

[35] W. Zhang, Fluorocarbon stationary phases for liquid chromatography applications, Journal of Fluorine Chemistry, 129 (2008) 910-919 https://doi.org/10.1016/j.jfluchem.2008.07.001.

[36] A. Glineur, K. Nott, P. Carbonnelle, S. Ronkart, G. Purcaro, Development And Validation Of A Method For Determining Estrogenic Compounds In Surface Water At The Ultra-Trace Level Required By The Eu Water Framework Directive Watch List, Journal of Chromatography A, (2020) 461242 https://doi.org/10.1016/j.chroma.2020.461242.

[37] S. Barreca, M. Busetto, L. Colzani, L. Clerici, D. Daverio, P. Dellavedova, S. Balzamo, E. Calabretta, V. Ubaldi, Determination of estrogenic endocrine disruptors in water at sub-ng L− 1 levels in compliance with Decision 2015/495/EU using offline-online solid phase extraction concentration coupled with high performance liquid chromatography-tandem mass spectrometry,

Microchemical Journal, 147 (2019) 1186-1191 https://doi.org/10.1016/j.microc.2019.04.030.

[38] B. Debrus, D. Guillarme, S. Rudaz, Improved quality-by-design compliant methodology for method development in reversed-phase liquid chromatography, Journal of pharmaceutical and biomedical analysis, 84 (2013) 215-223 https://doi.org/10.1016/j.jpba.2013.06.013.

[39] ICH Harmonised Tripartite Guideline, Pharmaceutical Development Q9, Quality risk management, (2005).

[40] K. Ishikawa, What is total quality control? The Japanese way, Prentice Hall1985.

[41] G.A. Lewis, D. Mathieu, R. Phan-Tan-Luu, Pharmaceutical experimental design, CRC press1998.

[42] A. Asghar, A.A. Abdul Raman, W.M.A.W. Daud, A comparison of central composite design and Taguchi method for optimizing Fenton process, The Scientific World Journal, 2014 (2014) 869120 https://doi.org/10.1155/2014/869120.

[43] D. Grover, Z. Zhang, J. Readman, J. Zhou, A comparison of three analytical techniques for the measurement of steroidal estrogens in environmental water samples, Talanta, 78 (2009) 1204-1210 https://doi.org/10.1016/j.talanta.2008.12.049.

[44] T. Anumol, S. Wu, M.M. Dos Santos, K.D. Daniels, S.A. Snyder, Rapid direct injection LC-MS/MS method for analysis of prioritized indicator compounds in wastewater effluent, Environmental Science: Water Research & Technology, 1 (2015) 632-643 https://doi.org/10.1039/C5EW00080G.

[45] C. Boix, M. Ibáñez, J.V. Sancho, J. Rambla, J.L. Aranda, S. Ballester, F. Hernández, Fast determination of 40 drugs in water using large volume direct injection liquid chromatography–tandem mass spectrometry, Talanta, 131 (2015) 719-727 https://doi.org/10.1016/j.talanta.2014.08.005.

[46] A. Garcia-Rodríguez, V. Matamoros, C. Fontàs, V. Salvado, The influence of Lemna sp. and Spirogyra sp. on the removal of pharmaceuticals and endocrine disruptors in treated wastewaters, International Journal of Environmental Science and Technology, 12 (2015) 2327-2338 https://doi.org/10.1007/s13762-014-0632-x.

5 Comparison of chemometric strategies for potential exposure marker discovery and false-positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of *Vaccinium* fruit supplements.

Section 5

Abstract

Untargeted liquid chromatographic-high-resolution mass spectrometric (LC-HRMS) metabolomics for potential exposure marker (PEM) discovery in nutrikinetic studies generates complex outputs. The correct selection of statistically significant PEMs is a crucial analytical step for understanding nutrition-health interactions. Hence, in this paper, different chemometric selection workflows for PEM discovery, using multivariate or univariate parametric or non-parametric data analyses, were comparatively tested and evaluated. The PEM selection protocols were applied to a small-sample-size untargeted LC-HRMS study of a longitudinal set of serum samples from 20 volunteers after a single intake of (poly)phenolic-rich Vaccinium myrtillus and Vaccinium corymbosum supplements. The non-parametric Games-Howell test identified a restricted group of significant features, thus minimizing the risk of false-positive retention. Among the forty-seven PEMs exhibiting a statistically significant postprandial kinetics, twelve were successfully annotated as purine pathway metabolites, benzoic and benzodiol metabolites, indole alkaloids, and organic and fatty acids, and five (i.e. octahydro-methyl-β-carboline-dicarboxylic acid, tetrahydro-methyl-β-carboline-dicarboxylic acid, citric acid, caprylic acid, and azelaic acid) were associated to Vaccinium berry consumption for the first time. The analysis of the area under the curve of the longitudinal dataset highlighted thirteen statistically significant PEMs discriminating the two interventions, including four intra-intervention relevant metabolites (i.e. abscisic acid glucuronide, catechol sulphate, methyl-catechol sulphate, and α-hydroxy-hippuric acid). Principal component analysis and sample classification through linear discriminant analysis performed on PEM maximum intensity confirmed the discriminating role of these PEMs.

Keywords: untargeted metabolomics; multivariate analysis; univariate parametric analysis; univariate non-parametric analysis; exposure markers; nutrikinetic

5.1   Introduction

Untargeted metabolomics has been extensively recognized as the leading approach for the investigation of potential exposure markers (PEMs) of food [1, 2], also in relation to food-health interaction studies [3] (i.e. nutrimetabolomics), as it allows in principle the comprehensive overview of the human metabolome. Untargeted metabolomics platforms usually consist in liquid chromatography (LC) hyphenated with high-resolution mass spectrometry (HRMS), which guarantees the determination of a much larger number of biological metabolites and/or a more comprehensive structural investigation, compared to other techniques such as gas chromatography coupled to HRMS or LC coupled to nuclear magnetic resonance [4]. The use of LC-HRMS platforms in nutrimetabolomics generally results in very complex outputs of data, which require the implementation of proper workflows for their pretreatment and processing [5]. Currently, no standard protocol has been defined to this aim, even though many workflows of data processing and analysis have been proposed [6]. Data handling is usually performed by chemometric tools, specifically selected according to the study design adopted, which may largely vary depending on a number of factors, such as study design (i.e cross over or parallel), the number of interventions, sample size, and data acquisition strategies (e.g. standalone full scan or coupled with MSn acquisitions) [5, 7]. In nutrimetabolomics studies, food PEMs discovery commonly involves the application of multivariate analysis, using partial least square-discriminant analysis (PLS-DA) and the variable importance in projection (VIP) classifier as filtering strategy of statistically significant features from the original dataset [8-10]. However, multivariate methods such as PLS-DA-VIP, due to their probabilistic nature, tend to select false positive (variables not causally related to groups) and/or false negative (missing of relevant features) [11]. On the other hand, the multiple univariate approaches, which have been applied less frequently in nutrimetabolomics, although not associated per se with false discovery risks, need the verification of some data distribution pre-requisites, i.e. normality or homoscedasticity [12].

However, it should be emphasized that multiple univariate approaches have been frequently applied as parametric methods (e.g. t-test and ANOVA), although the sample size was so small to suggest the absence of the aforementioned prerequisite [13, 14], thus involving a high risk of generating false positive. In order to reduce the false positive risk, an adjustment of the false positive rate should be performed for the application of multiple inference tests [15]. Conversely, the use of non-parametric tests is much less investigated [16], although the application of this type of chemometric tools would release the data analysis procedure from the constraints of homoscedasticity and normal distribution of data. Even though the impact of different chemometric tools in the PEMs selection is expected to be remarkable, this issue is poorly investigated in the literature, especially in LC-HRMS nutrimetabolomics studies that, as mentioned before, are prone by their nature to generate a large number of features to be tested for their statistical significance [17]. It is therefore important to compare the effects of the use of different chemometric approaches for the selection of statistically significant PEMs. This issue is particularly important in the light of the intrinsic complexity of the biological fluids (e.g. serum and urine) investigated in this type of studies. Among these matrices, serum and plasma are particularly relevant as matrices that contain substances responsible for a direct biological activity, the study of which is of remarkable importance to establish a cause-effect role between the intake of a certain food and the human health benefits.

Based on the aforementioned considerations, this paper aimed at comparing the effectiveness of chemometric protocols for PEMs discovery and false positive reduction, using different data analysis approaches, which included the widely adopted PLS-DA-VIP, the parametric t-test (before and after P-value adjustment), and the non-parametric Wilcoxon and Games-Howell tests (P-value adjusted for both tests), the latter never investigated before in untargeted nutrimetabolomics. The protocols have been tested on the untargeted LC-HRMS analysis of serum samples from an intervention study of *Vaccinium myrtillus* (VM) and *Vaccinium corymbosum* (VC) supplements as a representative application. In this regard, it

should be emphasized that this kind of application is of considerable intrinsic importance for nutrimetabolomics purposes. In fact, among fruits, small berries represent an abundant source of phenolic compounds in human diet [18]. In detail, the fruits belonging to the *Vaccinium* genus, above all *V. myrtillus berries*, have been suggested as functional foods and used for supplement and drug preparation [19, 20]. Moreover, in vitro studies have shown anti-proliferative and pro-apoptotic effects of polyphenol-rich berry extracts against different prostate cancer cell lines [21] and the chemopreventive properties of an anthocyanin-rich *V. myrtillus* extract were suggested in a pilot study with patients affected by colon cancer [22]. It should also be noted that few targeted [23, 24] and only one untargeted [1] metabolomics studies have been published so far on human serum and urinary (poly)phenolic metabolites of *Vaccinium* berries. **Figure 1** illustrates a comprehensive and intuitive scheme of the analytical workflow adopted in this study for the PEMs discovery in human serum after the administration of bilberry or blueberry supplement.
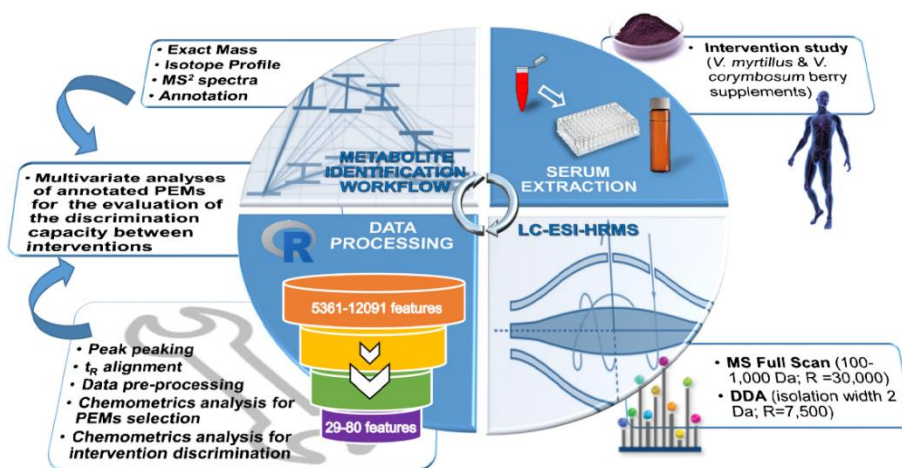

Figure 1 – Graphical workflow of the experimental steps followed in this study

5.2   Materials and Methods

5.2.1   *Reagents and standards*

Methanol and acetonitrile LC-MS Ultra CHROMASOLV™ and formic acid were purchased from Sigma Aldrich (St. Louis, MO, USA). The ultrapure water was obtained by purifying demineralized water in a Milli-Q system from Millipore (Bedford, MA, USA). O-hydroxy-hippuric acid, abscisic acid, azelaic acid, 1-methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic and L-tryptophan-2',4',5',6',7'-d5 (TRI-d5) reference standards were obtained from Sigma Aldrich, while trans-cinnamic-d5 acid (CIN-d5), N-Benzoyl-d5-glycine (Hippuric acid-d5, HIP-d5) from CDN ISOTOPES Inc.

5.2.2   *Vaccinium myrtillus and Vaccinium corymbosum berry supplements*

The supplements were obtained by cryo-milling VM and VC freeze-dried berries, as detailed in the section 1 of the *Supplementary material*. The supplements were characterized for total soluble polyphenols, total monomeric anthocyanins and for some phenolic compounds, using methods elsewhere described [25]. Full details of the characterization analysis (**Tables S1** and **S2**) and results obtained (**Tables S3-S5**) are reported in the section 2 of the *Supplementary material*.

5.2.3   *Study design*

The study was a randomized, single-blinded two-arms intervention of single intake of VM and VC, involving twenty healthy voluntary subjects (11 males and 9 females aged between 25 and 60 years). The study followed the guidelines set by the Helsinki Declaration (http://www.fda.gov/ohrms/dockets/dockets/06d0331/06D-0331-EC20-Attach-1.pdf) and all subjects provided written informed consent prior to the study. Ethical approval for freeze-dried VM and VC powdered supplement administration was obtained for a phase I-II study (approval n. SPE 14.178 AOUC, 30th May 2016). All the subjects convened early in the morning after 10 h of fasting and were randomly divided in two groups according to an electronic

randomisation key. A single dose of 25 g of VM supplement mixed with 500 mL of water was orally administered to the first group, whereas the same amount of VC berry supplement mixed with 500 mL of water was orally administered to the second group. This dose was chosen according to previous studies, as detailed in the section 3 of the Supplementary material. Serum was collected at baseline and different sampling times (30, 60, 120, 240 and 360 minutes) after the supplement consumption, as described in section 3 (**Table S6**) of the Supplementary material. After the collection, samples were divided in aliquots of 500 µL, frozen at -80ºC and stored until LC-HRMS analysis was performed.

5.2.4   Serum extraction

Serum extraction was performed on Waters Ostro 96-well plates as reported elsewhere [1]. Quality of the untargeted analysis was guaranteed by the use of quality control (QC) samples and internal and external standards. See details in the section 4 of the *Supplementary material*.

5.2.5   LC-MS and LC-MS/MS analysis

LC analysis was performed on a Dionex (Sunnyvale, CA, USA) Ultimate 3000 HPLC system equipped with a Kinetex C18 column (150 mm x 2.1 mm I.D., particle size 2.6 µm) and a guard column containing the same stationary phase (Phenomenex, Torrance, CA, USA). The LC system was coupled to a hybrid linear ion trap Fourier Transform (LTQ FT) Orbitrap high-resolution mass spectrometer (Thermo Fisher, Waltham, MA, USA) by an electrospray ionization (ESI) probe for MS and MS/MS analysis both in positive (PI) and negative (NI) ionization. Each sample was analysed under PI and NI, using two different mass acquisition methods for each ionization mode. The first method consisted of a Full Scan method (mass range from 100 to 1000 Da) at a mass resolution of 30,000 FWHM (m/z 400) in centroid mode. The second method included data dependent acquisition (DDA) scans for the three most abundant ions per scan at resolution of 7500 FWHM. Product ions were generated in the LTQ trap at collision energy

of 35 eV using an isolation width of 2 Da. Further details related to MS acquisition are reported in the section 5 of the *Supplementary material*.

5.2.6   Data pre-treatment

All raw data were manually inspected using the Qual browser module of Xcalibur version 2.0.7 (Thermo Fisher). The LC-MS raw files were converted to mzML format using the MSConvert utility included in ProteoWizard [26]. Then, the mzML files were processed with the XCMS R package that allows for obtaining detection and retention time alignment of features (i.e. ions characterized by a unique exact mass and retention time) [2, 27]. During this processing, the maximum value for mass accuracy and retention time deviation was set equal to 5 ppm and 2 seconds, respectively. Data pre-processing of Full Scan acquisitions was performed according to Garcia-Aloy et al. 2020 [28]. Briefly, each dataset was filtered to discard noisy and irrelevant features. Firstly, those features exhibiting a mass defect between 0.4 and 0.8 Da were removed from each data set. This procedure is based on the experimental evidence that exact masses mainly associated to metabolites of interest in this study (i.e. approximately 150-500 Da) have mass defects probabilistically distributed outside the aforementioned mass defect interval [29]. Subsequently, the signals that showed a higher variation coefficient in QC samples than within study samples were also excluded. A random number between 1 and 500 replaced all zero values. These thresholds were fixed empirically according to the intensity level representing the noise of the chromatogram. The principal component analysis (PCA) on log-transformed and Pareto-scaled data was applied to check data quality (i.e. possible batch effect) and to detect the presence of outliers.

5.2.7   Chemometric analyses

PEMs in postprandial responses were selected separately for VM and VC interventions using a selection protocol consisting in a two-step procedure: i) statistical significance among time points within a same intervention and ii) nutrikinetic discrimination between interventions.

5.2.7.1   Statistical significance among time points of a same intervention

For the identification of PEMs suitable for the discrimination among time points, the following approaches were investigated: univariate post-hoc non-parametric Games-Howell and Wilcoxon signed-rank tests, univariate parametric t-test, and multivariate PLS-DA-VIP. A short overview of the characteristics of these tests is reported in section 6 of the Supplementary material (**Table S7**).For the univariate approaches, we first selected those features with increasing mean values of chromatographic intensity in at least two consecutive time points, against baseline samples (i.e. time point "0 min"). Normal distribution of data was investigated by the application of Shapiro-Wilk's test for multiple variables, while homogeneity of variance was evaluated adopting a Bartlett's test for multiple independent variables, comparing the variances of VM and VC datasets in both polarities. Then, the t-test and the post-hoc non-parametric Games-Howell and Wilcoxon tests were applied to evaluate the statistical significance between time point "0 min" and all the others, setting the threshold P-value at 0.05. For both t-test and non-parametric tests, the Benjamini-Hochberg procedure was applied to decrease the false discovery rate (significance level 5%). The multivariate PLS-DA-VIP approach was performed on (i) the pre-treated VM and VC datasets and (ii) using the plsda package implemented with VIP scoring, setting the inclusion cut-off as > 1, as elsewhere reported [11, 16].

5.2.7.2   VM and VC discrimination by nutrikinetic data

For each feature selected as significant in VM and/or VC, the baseline sample was subtracted from the intensities of the other time points, and in case of negative values, the replacement by zero was performed. Nutrikinetic significant features were selected using a two-step procedure based on their nutrikinetic curve behaviour. For that purpose, according to a previous intervention study [28], features that in at least two consecutive points exhibited the 25th percentile of one group (n = 2) higher than the 75th percentile of the other group (n = 8), were included in the discrimination process based on comparison of area under the

curve (AUC). AUC of each selected feature was calculated between time 0 and 360 min using the pracma R package. Differences of AUCs among diets were tested using the Kruskal-Wallis test and the obtained P-values were adjusted using the Benjamini Hochberg method. Adjusted P-values < 0.05 were considered statistically significant. Descriptive analysis for evaluating the discrimination between the two interventions was carried out by means of PCA using the FactoMineR R package. Quality control of PCA was performed using QCs by visually verifying if PCA object scores obtained by replicated injections of the QC sample were close to the origin of PCA coordinates. Linear discriminant analysis (LDA) was performed on QC, VC, and VM sample groups made up by the annotated PEMs discriminating the two interventions, using the MASS R package.

5.2.8 *Metabolite identification*

The identification of features of interest, evidenced by data analysis, was performed according to the criteria previously reported by our research group [19], using the workflow described below. (i) Comparison of the exact mass of the experimental precursor ion with the pseudo-molecular ions proposed by the MzCloud (www.mzcloud.org), Humane Metabolome Database (HMDB, www.hmdb.ca), MassBank of North America (MoNA, http://mona.fiehnlab.ucdavis.edu), and Kyoto Encyclopedia of Genes and Genomes (KEGG) libraries, selecting a mass accuracy (Δ) ≤ 5 ppm as tolerance threshold. (ii) Export of the isotopic profile of pseudo-molecular ions selected within libraries at step (i) and subsequent comparison with the isotopic profile of the experimental precursor ions, selecting features providing an isotope ratio percentage difference of 20% as tolerance threshold. (iii) Structural elucidation of features of interest through the evaluation of MS/MS spectra obtained with DDA mass method in comparison with matched library mass spectra. (iv) Feature annotation performed according to golden standards for metabolomics [30].

5.3 Results and discussion

5.3.1 Assessment of data quality

Internal standards in study samples and QC samples showed a variation of retention time and mass accuracy below 2 seconds and 5 ppm, respectively, thus highlighting the correct data acquisition. Moreover, the variation of integrated area of the surrogate standards (added before the extraction) and internal standard (added before the analysis) in all the QC samples resulted lower than 20% confirming the repeatability of the extraction and excluding the possibility of signal drift and carry over phenomena during the LC-MS and LC-MS/MS analysis. Moreover, by means of Pareto-scaled PCA performed on raw data (data not shown), it was possible to observe that QC samples are well grouped in component space, showing no batch effect, or trend due to the injection order (i.e. drop in signal intensity). Study samples are distributed homogenously with no visible trends according to post-prandial time point.

5.3.2 Chemometric workflow for feature selection

After data pre-treatment, complex data sets were obtained for both interventions, being the total number of features 12091 and 5361 for PI and NI, respectively. The supervised multivariate PLS-DA-VIP analysis was performed on these groups of features, as commonly done in literature, highlighting a very high number of statistically significant features (2283-8639, depending on the dataset considered). This result is probably due to the tendency of the PLS-DA-VIP approach to select false positives. In fact, this criterion is very reasonable to discard irrelevant variables, but it may have drawbacks if used for assessing the significance of features [11]. Moreover, model classification error rate of maximized distances was very high, ranging from 0.9 to 0.7, regardless the number of selected components for PLS-DA modelling. For the application of univariate data analysis, the total number of features was preliminarily filtered by a selection procedure that preserves only the features showing increasing intensities in at least two consecutive time points. This feature selection resulted

in a number of relevant postprandial features ranging from 4245 to 4932, depending on the intervention and acquisition mode (**Table 1**). The obtained datasets were checked for (i) normal distribution and (ii) homogeneity of variance (i.e. homoscedasticity). Normality was evaluated by the application of the Shapiro-Wilk's test for multiple variables. Shapiro-Wilk's correlation coefficients and P-values (in bracket) ranged as follows: (i) VC – NI 0.538-0.897 ($1.0 \cdot 10$-5-$4.9 \cdot 10$-2); (ii) VC – PI 0.656-0.761 ($3.0 \cdot 10$-3-$4.9 \cdot 10$-2); (iii) VM – NI 0.366-0.807 ($1.0 \cdot 10$-7-$3.4 \cdot 10$-2); (iv) VM – PI 0.609-0.734 ($2.0 \cdot 10$-4-$5.0 \cdot 10$-3), thus highlighting the absence of normality of the investigated datasets. Homogeneity of variance was evaluated adopting the Bartlett's test for multiple independent variables, comparing the variances of VC and VM datasets in both polarities. The results obtained were characterized by P-value < 0.05 (confidence interval $= 0.95$), thus evidencing lack of homogeneity among variances of the aforementioned datasets. The aforementioned univariate tests were used for the PEMs selection within the aforementioned datasets of relevant postprandial features, obtaining more or less populated groups of statistically significant features in each acquisition polarity (**Table 1**). The adoption of the t-test generated the largest group of "statistically significant" PEMs, which surely includes a high number of false positives, even after the P-value correction using the Benjamini-Hochberg procedure (1051-2154 features, depending on the dataset considered), due to the non-normal and heteroscedastic distribution of data, which should discourage the use of parametric test for PEMs selection. The Wilcoxon test was found to be much more restrictive (75-359 features) than the adjusted t-test, in accordance with its non-parametric character, which does not require compliance with the conditions of normality and heteroscedasticity of data distribution, making it therefore more suitable for the treatment of small sample size, for which the above-mentioned conditions usually do not occur. A significantly reduced number of statistically significant features (i.e. 80 and 29 for VM and VC interventions, respectively) was identified using the Games-Howell test, thus highlighting its strong selectivity, which greatly reduces the probability of including false

positives in the PEMs group. It is worth noting that the group of features selected by the Games-Howell test represents a subset included in all the groups of features selected by the other tests investigated. Moreover, using as representative cases some features that can be easily annotated due to their previous identifications in bilberry and/or blueberry intervention studies, it is interesting to evaluate how these features are treated by the procedures of PEMs selection here tested. For instance, benzoic acid was annotated here (Δ=-0.4 ppm) after the intake of VC supplement and retained in the statistically significant PEMs group using the adjusted t-test. However, this feature was excluded when PLS-DA-VIP or the non-parametric Wilcoxon and Games-Howell tests for PEMs selection were applied. Moreover, hippuric acid was annotated here (Δ = -0.6 ppm) in both interventions and considered statistically significant by PLS-DA-VIP, adjusted t-test, and Wilcoxon test but excluded by the Games-Howell treatment. It is interesting to note that these features were considered statistically significant in a study investigating plasma after blueberry intake using ANOVA and post-hoc Bonferroni correction [24]. As further examples, four metabolites (i.e. hydroxy-methoxy hippuric acid, hydroxy-(hydroxy-methoxyphenyl)-pentenoic acid glucuronide, dihydroxyphenyl propionic acid glucuronide, and hydroxypheny propionic acid sulphate) recently annotated as PEMs of bilberry intake [1], resulted here statistically significant by applying the Wilcoxon signed-rank test, while they were discarded by the Games-Howell test, due to the different assumptions made with respect to Wilcoxon test. All these examples evidenced that the selection of significant features is strongly dependent on the statistical tool adopted. Based on these findings, great caution should be used in reporting metabolites as statistically significant for a given intervention. The application of highly conservative statistical methods, i.e. prone to minimizing the risk of false positive results is certainly useful in this sense and should be encouraged, although it involves a certain risk of excluding false negative from the significant dataset. On the contrary, the use of less restrictive tests should be used with great caution in PEMs discovery, even though they give a wider overview of features that could

become significant when moving from a pilot study, i.e. characterized by a limited number of subjects, to one on a larger validation cohort.

Table 1 – Number of relevant postprandial features and potential exposure markers (PEMs) in each dataset (intervention and acquisition polarity) identified using the selection strategies based on univariate non-parametric adjusted post hoc Games-Howell test, pairwise Wilcoxon signed-rank test, and the univariate parametric t test, before and after the Benjamini-Hochberg adjustment procedure (BH-adj). VM = Vaccinium myrtillus supplement intervention; VC = Vaccinium corymbosum supplement intervention.

Dataset	Relevant postprandial features	t-test	BH-adj	Wilcoxon	Games-Howell
VM ESI (+)	4731	4682	1957	359	29
VC ESI (+)	4932	4395	2154	146	18
VM ESI (−)	4655	3286	1075	224	51
VC ESI (−)	4245	2582	1051	75	11

5.3.3 PEMs annotation

Among the features selected by the Games-Howell test, corresponding to 47 PEMs, the annotation protocol allowed for identifying 5 features in PI and 13 features in NI, corresponding to a total of 12 PEMs. **Table 2** reports the annotated metabolites, providing their spectral characteristics, formula, time points for statistically significant maximum intensities (T_{max}) with related P-values, the annotation level, and for annotation levels I and II, the HMDB/KEGG identification numbers. Among the twelve PEMs identified, eight were annotated only after VM intervention (peaks 2-6, 9, 11, and 12), three only after VC ingestion (peaks 7, 8, and 10), and only one (peak 1) as common response to the consumption of the two supplements. Peak 1 showed two pseudo-molecular ions in NI Full Scan spectra, i.e. at m/z 335.0504 and m/z 167.0214. The second ion was characterized by a fragmentation pattern that matched with uric acid, as reported in mzCloud spectral library and literature findings [31]. Peak 2 was putatively annotated as citric acid, due the occurrence of precursor ions at m/z 193.0335 and 191.0202, in PI and NI, respectively. MS/MS fragmentation pattern included as main fragments in NI mode peaks at m/z 173 (loss of a water molecule)

and m/z 111.0091 (loss of acetic acid), whilst in PI m/z 132.1019 (loss of acetic acid) was highlighted. This attribution was confirmed by mzCloud spectral library and previous researches [32]. Peak 3 exhibited $[M–H]^-$ ion at m/z 267.0744, characterized by the neutral loss of 132 Da (pentose moiety) that led to the formation of a fragment at m/z 135.0313 associated to hypoxanthine, in accordance with spectral libraries and previous findings [33]. Thus, this metabolite was putatively identified as inosine, a nucleoside naturally occurring in the pathway of purine metabolism after vegetable and/or fruit intake [34]. Peak 4 gave rise to a $[M–H]^-$ pseudo-molecular ion at m/z 194.0454 Da, which fragmented by losing 44 Da (CO_2) originating the fragment at m/z 150.0561. Since the pseudo-molecular ion matched with high accuracy (Δ=-2.6 ppm) the mass of a hydroxy-hippuric acid and no other fragments were detected in the MS^2 spectrum, α-hydroxy-hippuric acid, p-hydroxy-hippuric acid, and o-hydroxy-hippuric acid reference standards were injected, obtaining tR of 3.3, 3.8, and 5.4 minutes, respectively. Accordingly, peak 4 was annotated at level I as α-hydroxy-hippuric acid. Peak 5 was putatively identified as benzodiol (catechol) sulphate owing to the presence of the typical loss of sulphate (m/z 79.9576) and the appearance of the hydroxy-phenol ion at m/z 109.0297 in the MS/MS spectrum [1, 35]. An analogous fragmentation pattern was observed for peak 6, putatively identified as methyl-catechol sulphate, due to the presence in the MS/MS spectrum of a peak at m/z 188.9860 (loss of methyl), together with typical losses of sulfonic (m/z 123.0454) and sulphate (m/z 108.0220) groups. Peak 7 (PI: m/z 279.1329, NI: m/z 277.1185) and peak 8 (PI: m/z 275.1024, NI: m/z 273.0875) were recognized as indole alkaloids derivatives, based on their mass accuracy and similar MSn fragmentation patterns (**Figure S1** of the *Supplementary material*). Spectra of both peaks in PI (**Fig. S1**, box A and box E) were characterized by the neutral loss of C2H3NO2 (73 Da, iminoacetic acid), typical of carboxylic acid derivatives of carbolines [36]. Moreover, under NI mode, the pseudo-molecular ions (i.e. m/z 277.1185 and 273.0875) underwent the sequential loss of two CO_2 moieties, indicating the presence of two carboxylic groups (**Fig. S1**, box B-C and

box F-G). In order to confirm the attribution of peaks 7 and 8 to the carbolines class, their fragmentation patterns were compared with the one of 1-methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid reference standard (**Fig. S1**, boxes D-H). Interestingly, the pseudo-molecular ion of peak 8 in NI mode (*m/z* 273.0875) lost a carboxylic group, originating the feature at *m/z* 229.0982 (**Fig. S1**, box F), the fragmentation pattern of which (**Fig. S1**, box G) matched the one of the analytical standard 1-methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid (**Fig. S1**, box D) with good mass accuracy and similar ions intensities. It should also be noted that peak 7 showed in both PI (**Fig. S1**, boxes A vs. E) and NI (**Fig. S1**, boxes B vs. F and C vs. G) MS^n spectra a constant difference of about 4.03 Da compared to corresponding ions of peak 8. Hence, peak 7 was annotated as octahydro-methyl-β-carboline-dicarboxylic acid. Peak 9 showed the $[M–H]^-$ pseudo-molecular ion in NI at *m/z* 439.1599, which fragmented through the typical neutral loss of glucuronide-conjugated compounds (176 Da) giving rise to the product ion at *m/z* 263.1288. This ion further fragmented originating the *m/z* 219.1391 and 153.0921 ions (**Figure S2-A** of the *Supplementary material*), which can be attributed to abscisic acid, as also highlighted by injecting the abscisic acid reference standard (**Figure S2-B**). Peaks 10 and 11 were annotated as caprylic acid and azelaic acid, respectively, by comparing their mass spectra with those present in spectral databases (i.e. mzCloud and HMDB, see **Table 2**). In addition, azelaic acid was further confirmed at Level I by injecting the analytical standard. Finally, peak 12 was putatively annotated as hydroxy-phenyl propionic acid sulphate thanks to the mass accuracy of its $[M–H]^-$ pseudo-molecular ion (*m/z* 245.0119, Δ=-2.4 ppm), the comparison with MS^2 spectral libraries, and based on findings of previous clinical studies on VM intake [1].

Table 2 – List of metabolites found in serum samples after acute ingestion of V. myrtillus (VM) and V. corymbosum (VC) supplements. Retention time (tR, min), experimental mass in Full Scan spectrum (Da), main MS/MS fragments (base peak in bold), proposed formula, mass accuracy (Δ, ppm), time point of max intensity (Tmax, min) and P-value within interventions (in brackets), and HMDB/KEGG identification numbers. Roman numbers refer to the level of annotation: (I) spectra and tR matched with reference standard; (II) spectra matched with reference libraries; (III) spectra matched with literature information. n.a.= not available.

Peak	t_R	Full Scan	MS/MS	Formula	T_{max} (*P*-value)	Identification	HMDB/KEGG
1	1.2	(-) 335.0504 [2M-H]- (-) 167.0214	**167.0214** **124.0156**, 96.0207	$C_5H_4N_4O_3$	VM: 60 min (0.046) VC: 30 min (0.037)	Uric acid [II]	HMDB0000289/C00366
2	1.2	(+) 193.0335 [M+H]+ (-) 191.0202 [M-H]-	**132.1019** 173.0093, **111.0091**	$C_6H_8O_7$	VM: 30 min (0.041) VM: 60 min (0.010)	Citric acid [II]	HMDB0000094/C00158
3	1.5	(-) 267.0744 [M-H]-	**135.0313**	$C_{10}H_{12}N_4O_5$	VM: 60 min (0.013)	Inosine [II]	HMDB0000195/C00294
4	3.1	(-) 194.0454 [M-H]-	**150.0561**	$C_9H_9NO_4$	VM: 240 min (0.0001)	α-Hydroxy-hippuric acid [I]	HMDB0013678 HMDB0006116
5	3.9	(-) 188.9856 [M-H]-	**109.0297**	$C_6H_6O_5S$	VM: 360 min (0.008)	Benzendiol (catechol) sulphate [II]	HMDB0059724
6	4.7	(-) 203.0014 [M-H]-	188.9860, **123.0454**, 108.0220	$C_7H_8O_5S$	VM: 360 min (0.028)	Methyl-catechol sulphate [II]	HMDB0240663
7	5.23	(+) 279.1329 [M+H]+ (-) 277.1185[M-H]- (-) 233.1232	235.1376, 206.1110, **163.1165** **233.1232** **189.1336**, 187.1179, 146.0914, 120.0759, 96.0736	$C_{14}H_{18}N_2O_4$	VC: 60 min (0.023(+), 0.035(-))	Octahydro-methyl-β-carboline-dicarboxylic acid [III]	n.a.
8	5.30	(+) 275.1024 [M+H]+ (+) 202.0861 (-) 273.0875 [M-H]- (+) 229.0982	231.1128, 202.0861, **159.0916** **184.0755**, 156.0802 **229.0982** **185.1084**, 183.0928, 142.0663, 116.0508, 92.0508,	$C_{14}H_{14}N_2O_4$	VC: 60 min (0.007(+), 0.008(-))	Tetrahydro-methyl-β-carboline-dicarboxylic acid [II]	HMDB0035115
9	5.9	(+) 265.1431 (+) 247.1326 (-) 439.1599 [M-H]-	**247.1329**, 199.1023 229.1431, 211.1327 **263.1288**, 219.1395, 175.0248, 153.0922	$C_{21}H_{28}O_{10}$	VM: 120 min (0.044)	Abscisic acid glucuronide [II]	HMDB0036093 HMDB0035140
10	6.3	(+) 145.1220 [M+H]+	**127.1130**, 109.1008	$C_8H_{16}O_2$	VC: 30 min (0.021)	Caprylic acid (hydroxy-octanone) [II]	HMDB00482/ HMDB0062511 HMDB62511/C06423
11	6.4	(-) 187.0980 [M-H]-	**125.0973**	$C_9H_{16}O_4$	VM: 360 min (0.028)	Azelaic acid [I]	HMDB00784/C08261
12	6.5	(-) 245.0119 [M-H]-	**165.0556**	$C_9H_{10}O_6S$	VM: 120 min (0.038)	Hydroxyphenyl propionic acid sulphate [II]	HMDB0094710

Bold MS/MS fragments refer to spectra base peaks.

5.3.4 Nutrimetabolomics significance

The twelve annotated metabolites reported in **Table 2** were categorized as (i) purine pathway metabolites, (ii) benzoic and benzodiol metabolites, (iii) indole alkaloids, and (iv) organic and fatty acids, according to their chemical class and metabolism. **Figures 2-5** report the serum postprandial kinetics of the 12 annotated PEMs, expressed as boxplots of the precursor ion intensity as a function of time. Each boxplot represents the interquartile range (75% of the intensity values are less than or equal to the top value of the box and 25% of the intensity values are less than or equal to the bottom value of the box), the upper and lower whiskers refer to the maximum and minimum data point, respectively, and the line within the box represents the median of the data.

5.3.4.1 Purine pathway metabolites

Figure 2 illustrates the kinetic profile of the pseudo-molecular ions of uric acid (peak 1) in VM (**Fig. 2A**) and VC (**Fig. 2B**), and inosine (peak 3) in VM (**Fig. 2C**). According to literature, these metabolites are correlated with the ingestion of fructose, rather than phenolic compounds, since fructose induces acute depletion of ATP and inorganic phosphate and causes increased activity of the enzymes involved in the degradation of purine nucleotides to inosine, hypoxanthine, xanthine, and finally uric acid [37]. Statistically significant variations ($P < 0.05$) of uric acid and inosine were observed at T_{max} 30-60 min, suggesting that the fructose-induced response increases uric acid concentration in plasma in the early stages of metabolism. The increase in uric acid is responsible of the raise of plasma antioxidant activity widely observed in literature after ingestion of fructose-rich fruits [34, 38]. The fast depletion of these metabolites after two-three hours from food intake is in agreement with the post-prandial trend commonly observed for phase II metabolites of methylxanthines, which are rapidly excreted through the urinary trait within eight hours from the intake [39, 40].
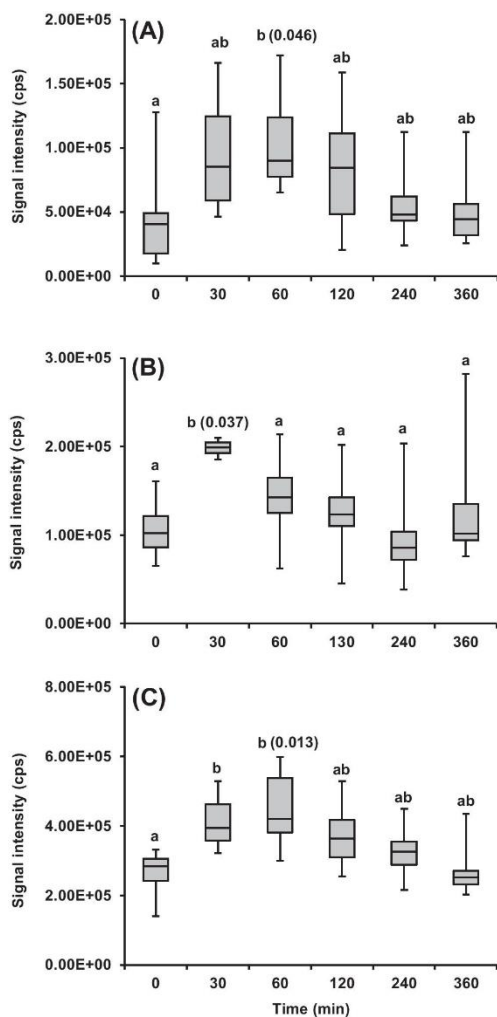
Figure 2 – Kinetic profiles of the pseudo-molecular ions (negative ionization mode) at *m/z* A 335.0504 (uric acid in VM), B 167.0214 (uric acid in VC), and C 267.0744 (inosine in VM). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ($P < 0.05$). T_{max} of each pseudo-molecular ion is labelled in brackets with its corresponding P value.

5.3.4.2 *Benzoic and benzodiol metabolites*

Figure 3 shows the kinetic profiles of α-hydroxy-hippuric acid (peak 4, **Fig. 3A**), the sulphate derivatives of catechol (peak 5, **Fig. 3B**) and methyl-catechol (peak 6, **Fig. 3C**), and hydroxyphenyl propionic acid sulphate (peak 12, **Fig. 3D**), which were detected only in response to the VM intervention. Differently from purine derivatives, the trend observed for these compounds reflected a diverse

postprandial scenario. In fact, the increase in signals intensities became statistically significant (P < 0.05) starting from two hours after the supplement ingestion (T_{max} between 120 and 360 min), suggesting that their occurrence in plasma is associated to microbiota activity, which mediates the degradation of polyphenols in smaller and more soluble molecules [41]. In detail, the occurrence in human plasma of both hippuric acid and catechol derivatives was already reported in literature [1, 23, 24] as potential markers of acute and/or long-term ingestion of bilberry and blueberry, and their occurrence related to quercetin glycosides, which are dominant flavonols in these fruits [1, 42, 43]. Conversely, phenyl-propionic acid derivatives may derive from caffeic acid degradation and/or from anthocyanins after B-ring cleavage, and C-ring opening and oxidation [44].
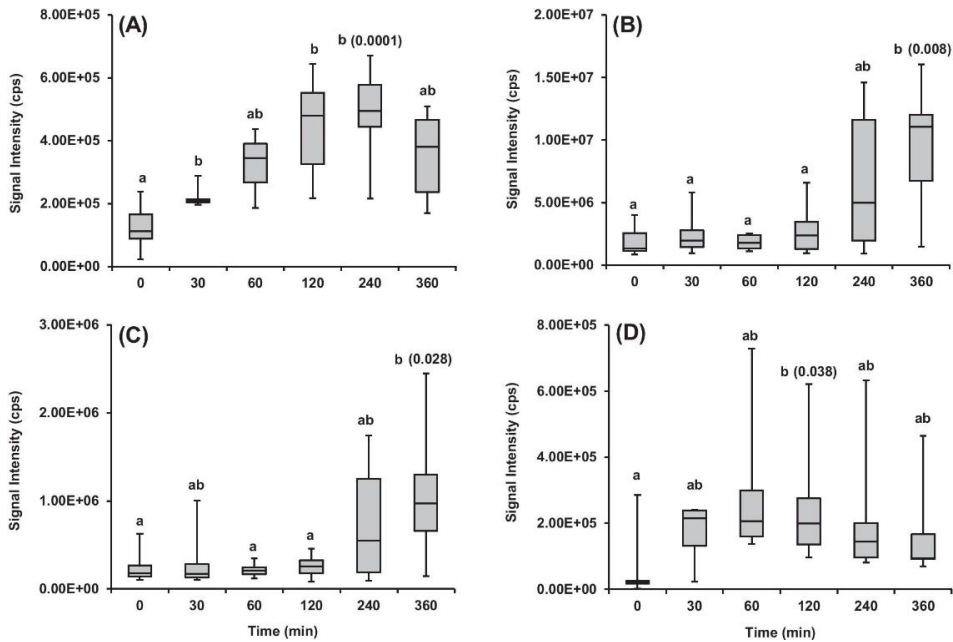
Figure 3 – Kinetic profiles of the pseudo-molecular ions (negative ionization mode) at *m/z* A 194.0454 (α-hydroxy-hippuric acid in VM), B 188.9856 (catechol sulphate in VM), C 203.0014 (methyl-catechol sulphate in VM), and D 245.0119 (hydroxyphenyl propionic acid sulphate in VM). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test (P < 0.05). Tmax of each pseudo-molecular ion is labelled in brackets with its corresponding P value.

5.3.4.3   Indole alkaloids

Figure 4 reports the kinetic profiles of the octahydro-methyl-β-carboline-dicarboxylic acid (peak 7, Fig. **4A-B**) and the tetrahydro-methyl-β-carboline-dicarboxylic acid (peak 8, **Fig. 4C-D**), which have been annotated here for the first time in human plasma in response to the VC intervention. Interestingly, both PEMs showed a significant increase in their intensities ($P < 0.05$) in plasma at T_{max} 60 min, then decreasing in the following time points, suggesting that they are rapidly absorbed by human organism and likewise excreted through the urinary trait, as highlighted elsewhere [45]. This group of chemicals is already known to occur in fruits, including berries, as well as in food processed products, suggesting that diet may directly contribute to their presence in human bio-fluids [36, 46, 47]. However, the occurrence of these alkaloids should not be necessarily associated to the intake of fruit, since other foods such as beer, coffee, cereal products, and vegetables contain these alkaloids and may therefore contribute to their presence in human biofluids [36, 45].

5.3.4.4   Organic and fatty acids

Figure 5 illustrates the kinetic profiles observed in response to the VM intervention for citric acid (peak 2, **Fig. 5A-B**), abscisic acid glucuronide (peak 9, **Fig. 5C-D**), and azelaic acid (peak 11, **Fig. 5F**). Moreover, the trend of caprylic acid (peak 10, **Fig. 5E**) after VC intervention is also shown. Citric acid showed a postprandial maximum intensity in between 30 and 60 min. Even though this compound is already known as one of the main organic acids occurring in various fruits, including blueberry and especially bilberry [48-50], these data represent the first evidence of a correlation between fruit intake and the occurrence of a statistically significant increase of citric acid in plasma. Abscisic acid glucuronide reached a T_{max} at 120 min after the supplement consumption, suggesting that it is involved in the phase II metabolism, even though the possible endogenous contribution associated with supplement intake cannot be excluded.

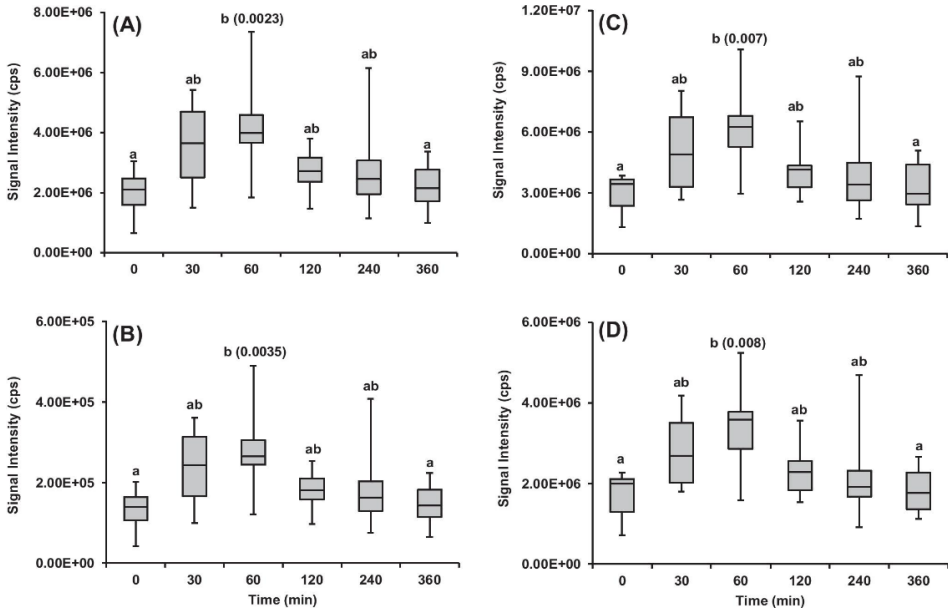**Figure 4** – Kinetic profiles of the pseudo-molecular ions (positive and negative ionization, respectively) at m/z A 279.1329 and B 277.1185 (octahydro-methyl-β-carboline-dicarboxylic acid in VC), C 275.1024 and D 273.0875 (tetrahydro-methyl-β-carboline-dicarboxylic acid in VC). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ($P < 0.05$). Tmax of each pseudo-molecular ion is labelled in brackets with its corresponding P value.

In fact, abscisic acid has been found at relevant concentrations in bilberry fruits as a well-known plant hormone involved in the regulation of biosynthesis of polyphenols, such as anthocyanins [51]. Caprylic acid and azelaic acid showed a postprandial maximum intensity at very different T_{max}, i.e. 30 min and 360 min after VC and VM interventions, respectively. In detail, caprylic acid, after the very fast absorption with an early post-prandial maximum intensity, exhibited a rapid decrease in the following time points, ascribable to its excretion by urine. On the other hand, the late intensity increase of the azelaic acid kinetic profile suggests a closer dependence on the phase I metabolism than the phase II one. The presence of fatty acids in serum has been poorly investigated in association with phenolic-rich fruit interventions. However, caprylic acid has been detected in human serum after pomegranate ingestion, probably due to its occurrence in seeds [52]. Moreover, fatty acids other than caprylic and azelaic have been detected in human

serum after the intake of mulberry and attributed to their occurrence in fruit seeds and peels [53]. Hence, to our knowledge, caprylic acid and azelaic acid have been reported here as PEMs of VM and VC interventions for the first time.
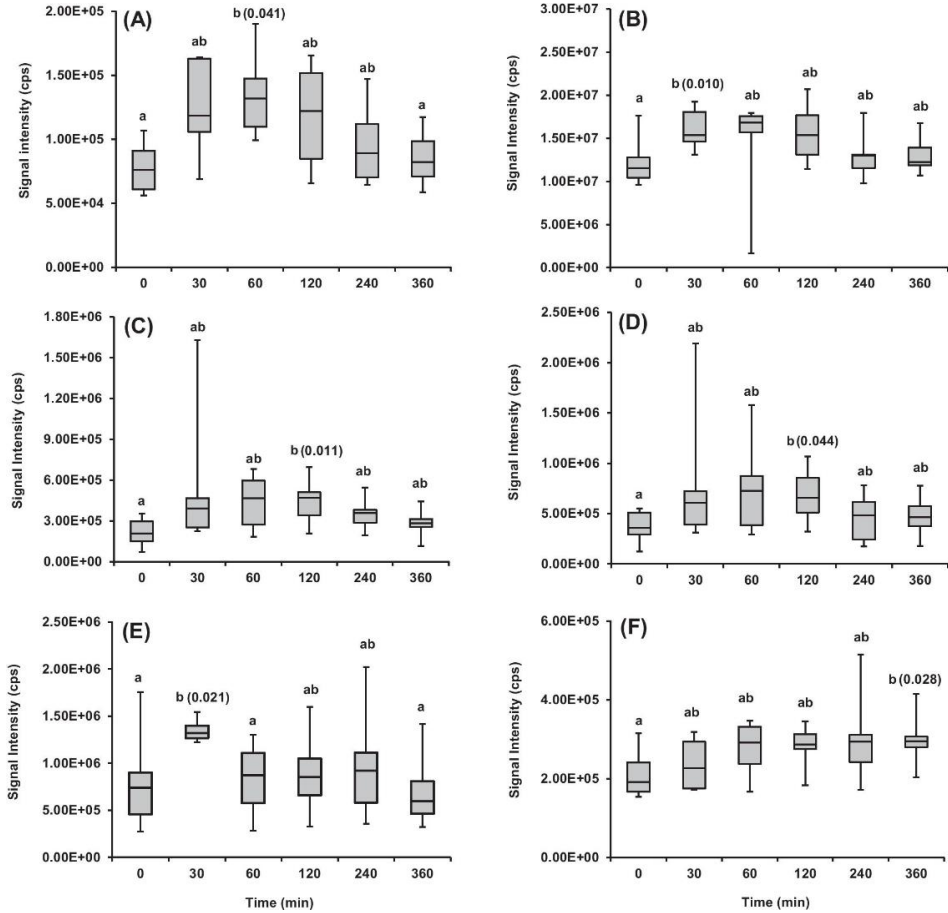
Figure 5 – Kinetic profiles of the pseudo-molecular ions at *m/z* A 191.0202 and B 193.0335 (citric acid in VM under negative and positive ionization, respectively), C 265.1431 and D 439.1599 (abscisic acid glucuronide in VM under positive and negative ionization, respectively), E 145.1220 (caprylic acid in VC under positive ionization), and F 187.0980 (azelaic acid in VM under negative ionization). Boxplots exhibiting different letters are significantly different according to the Games-Howell comparison test ($P < 0.05$). T_{max} of each pseudo-molecular ion is labelled in brackets with its corresponding P value.

5.3.5   *Discrimination between VM and VC interventions*

Comparison of serum metabolome profiles between the two berry interventions (performed according to the protocol described in section 2.7.2) showed ten and fifteen discriminant features in PI and NI, respectively, which correspond to thirteen PEMs. Among these, four metabolites matched the previously annotated PEMs, i.e. abscisic acid glucuronide, α-hydroxy-hippuric acid, catechol sulphate, and methyl-catechol sulphate, which were found to be statistically significant in VM intervention (**Table 2**). In order to provide a summarising description of the discrimination capacity between the two interventions by the thirteen PEMs, it is interesting to perform a multivariate analysis by means of PCA and LDA of the autoscaled intensity data of these PEMs, expressed as the highest intensity value determined for each feature under NI in the kinetic profile of each intervention. Moreover, the intensities of these PEMs in QC samples were included in both PCA and LDA, in order to make more informative the multivariate analyses, also avoiding misclassifications. Five principal components (PCs) characterized by eigenvalues > 1 and accounting for percentages of explained variance (EV) of 29.9%, 12.9%, 12.2%, 11.2%, and 8.2% were obtained (total EV=74.6%). **Table S8** in section 6 of the *Supplementary material* illustrates the loading values of the five significant PCs. Most features mainly contributed to only one PC with high absolute values of the loadings, thus highlighting a good separation of the original variables among PCs. For example, abscisic acid glucuronide, α-hydroxy-hippuric acid, catechol sulphate, methyl-catechol sulphate, and the unknown at *m/z* 287.0224 contributed significantly only to PC1, which alone accounts for about a third of the total variance of the original data. A detailed description of the contribution of the various features to the five significant PCs is reported in section 6 of the *Supplementary material*. **Figure 6** illustrates the score plots of PC1 vs. PC2, PC3, PC4, and PC5, which highlights a quite good separation of QC, VC, and VM samples in all the four PC spaces. Moreover, QCs were very close to the origin of the coordinates, suggesting a high accuracy and precision of the entire analytical procedure. LDA has been also carried out on the dataset of

the discriminating PEMs in QC, VC, and VM samples, in order to verify the correct assignment of each sample to its own group (i.e. QC, VM, and VC). The summary of the results obtained for the fitted and cross-validated (leave-one-out method) LDA classification are illustrated in **Table 3**. QC samples were correctly classified in all cases, both in fitting and in cross-validation, thus evidencing the robustness of this classification and confirming the results of PCA regarding the reliability of the analytical procedure here adopted. The LDA classification exhibited a very high robustness also for the VM and VC groups, since all samples were correctly classified in fitting and only two (i.e. one for each intervention) were misclassified in cross-validation. It should however be remarked that the two misclassified samples were attributed to the QC group, without any misleading attribution to the other volunteer group.
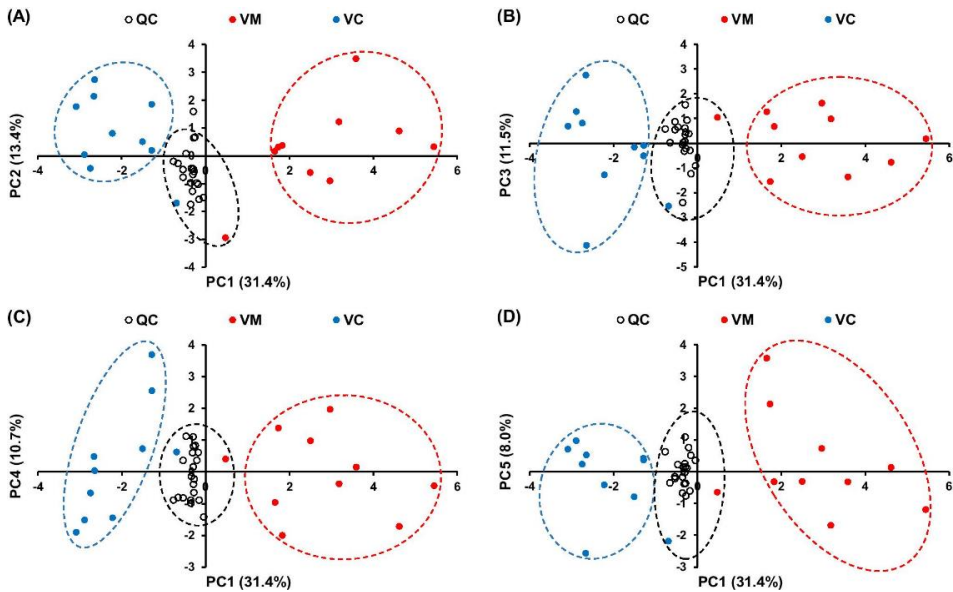
Figure 6 – Score plots of A PC1 vs. PC2, B PC1 vs. PC3, C PC1 vs. PC4, and D PC1 vs. PC5. PCA plots were obtained using the autoscaled intensity data of the PEMs discriminating the two interventions in serum samples from each volunteer, expressed as their features measured under NI at the maximum of their kinetic profile in each intervention. Ellipses identify samples assigned to the QC, VM, and VC groups by applying linear discriminant analysis of maximum intensity of each annotated PEMs after cross validation with the leave-one-out method.

Table 3 – Fitting (FIT) and cross-validation (CV) results of the Linear Discriminant Analysis based on the values of intensity data of the thirteen discriminating PEMs in serum samples from each volunteer, expressed as their features measured at the maximum of their kinetic profile in each intervention. Groups: quality control (QC), intervention with *Vaccinium corymbosum* supplement (VC), and intervention with *Vaccinium myrtillus* supplement (VM).

Put into Group		True Group		
		QC	VC	VM
QC	FIT	20	0	0
	CV	20	1	1
VC	FIT	0	10	0
	CV	0	9	0
VM	FIT	0	0	10
	CV	0	0	9
Total Number of Samples		20	10	10
Correct Attributions	FIT	20	10	10
	CV	20	9	9
Percentage of Correct Attribution	FIT	100%	100%	100%
	CV	100%	90%	90%

5.4 Conclusions

The comparative evaluation of different multivariate and univariate methods for PEMs selection evidenced that the number of significant features is strongly dependent on the statistical tool adopted, thus highlighting the importance of testing multiple statistical approaches for PEMs discovery. The protocol of untargeted metabolomics analysis and chemometric data treatment adopting the non-parametric Games-Howell test for PEMs selection within interventions, was suitable for selecting a restricted set of statistically significant features, excluding a number of false positive, which were conversely retained in the highly populated group of "significant features", generated by other statistical methods, commonly adopted in literature. Accordingly, this chemometric strategy could be fruitfully extended to the data treatment of intervention studies characterized by a similar study design (i.e. small sample size, randomized, single-blinded, parallel studies). In fact, it is more useful to identify a smaller set of PEMs, accepting a certain risk of excluding from this group some significant metabolites (false negatives), rather than selecting a larger group of PEMs, which however contains a high number of false positive. It should also be noted that with the presented approach it is possible to concentrate the annotation effort on a smaller but highly discriminant

and certainly significant set of PEMs. Twelve PEMs related to the intake of bilberry and/or blueberry supplements were annotated in serum samples, five of them (i.e. octahydro-methyl-β-carboline-dicarboxylic acid and tetrahydro-methyl-β-carboline-dicarboxylic acid for VC, citric acid for VM, and caprylic acid and azelaic acid for both VC and VM) reported here for the first time as serum metabolites of these interventions. Based on AUC analysis, thirteen PEMs were found statistically significant for the discrimination of VM and VC interventions. In this regard, it is remarkable that this group of PEMs included four intra-intervention relevant metabolites (i.e. abscisic acid glucuronide, catechol sulphate, methyl-catechol sulphate, and α-hydroxy-hippuric acid), thus confirming the validity of the selection workflow based on the Games-Howell test. PCA data exploration and LDA samples classification performed on PEMs intensity at T_{max} corroborating the discriminating role of the thirteen PEMs.

References

[1] C. Ancillotti, M. Ulaszewska, F. Mattivi, M. Del Bubba, Untargeted metabolomics analytical strategy based on liquid chromatography/electrospray ionization linear ion trap quadrupole/orbitrap mass spectrometry for discovering new polyphenol metabolites in human biofluids after acute ingestion of vaccinium myrtillus berry supplement, J. Am. Soc. Mass. Spectrom., 30 (2019) 381-402 https://doi.org/10.1007/s13361-018-2111-y.
[2] M. Ulaszewska, M. Garcia-Aloy, N. Vázquez-Manjarrez, M. Soria-Florido, R. Llorach, F. Mattivi, C. Manach, Food intake biomarkers for berries and grapes, Genes Nutr, 15 (2020) 1-35 https://doi.org/10.1186/s12263-020-00675-z.
[3] A. Koutsos, S. Riccadonna, M.M. Ulaszewska, P. Franceschi, K. Trošt, A. Galvin, T. Braune, F. Fava, D. Perenzoni, F. Mattivi, Two apples a day lower serum cholesterol and improve cardiometabolic biomarkers in mildly hypercholesterolemic adults: a randomized, controlled, crossover trial, Am. J. Clin. Nutr., 111 (2020) 307-318 https://doi.org/10.1093/ajcn/nqz282.
[4] L. Lacalle-Bergeron, D. Izquierdo-Sandoval, J.V. Sancho, F.J. López, F. Hernández, T. Portolés, Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio) markers, TrAC, Trends Anal. Chem., (2020) 116161 https://doi.org/10.1016/j.trac.2020.116161.
[5] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L. Brennan, C. Brunius, A. Bub, F. Capozzi, Nutrimetabolomics: an integrative action for metabolomic analyses in human

nutritional studies, Mol. Nutr. Food Res., 63 (2019) 1800384 https://doi.org/10.1002/mnfr.201800384.

[6] J. Stanstrup, C.D. Broeckling, R. Helmus, N. Hoffmann, E. Mathé, T. Naake, L. Nicolotti, K. Peters, J. Rainer, R.M. Salek, T. Schulze, E.L. Schymanski, M.A. Stravs, E.A. Thévenot, H. Treutler, R.J.M. Weber, E. Willighagen, M. Witting, S. Neumann, The metaRbolomics Toolbox in Bioconductor and beyond, Metabolites, 9 (2019) 200 https://doi.org/10.3390/metabo9100200.

[7] M.M. Hendriks, F.A. van Eeuwijk, R.H. Jellema, J.A. Westerhuis, T.H. Reijmers, H.C. Hoefsloot, A.K. Smilde, Data-processing strategies for metabolomics studies, TrAC, Trends Anal. Chem., 30 (2011) 1685-1698 https://doi.org/10.1016/j.trac.2011.04.019.

[8] C.S. Cuparencu, M.-B.S. Andersen, G. Gürdeniz, S.S. Schou, M.W. Mortensen, A. Raben, A. Astrup, L.O. Dragsted, Identification of urinary biomarkers after consumption of sea buckthorn and strawberry, by untargeted LC–MS metabolomics: a meal study in adult men, Metabolomics, 12 (2016) 31 https://doi.org/10.1007/s11306-015-0934-0.

[9] T. Xu, G. Feng, B. Zhao, J. Zhao, Z. Pi, S. Liu, F. Song, Z. Liu, A non-target urinary and serum metabolomics strategy reveals therapeutical mechanism of Radix Astragali on adjuvant-induced arthritis rats, J. Chromatogr. B, 1048 (2017) 94-101 https://doi.org/10.1016/j.jchromb.2017.01.040.

[10] L. Dickson, M. Tenon, L. Svilar, P. Fança-Berthon, J.-C. Martin, H. Rogez, F. Vaillant, Genipap (Genipa americana L.) juice intake biomarkers after medium-term consumption, Food Res. Int., 137 (2020) 109375 https://doi.org/10.1016/j.foodres.2020.109375.

[11] M. Cocchi, A. Biancolillo, F. Marini, Chemometric methods for classification and feature selection, Comprehensive Analytical Chemistry, Elsevier2018, pp. 265-299 https://doi.org/10.1016/bs.coac.2018.08.006.

[12] M. Vinaixa, S. Samino, I. Saez, J. Duran, J.J. Guinovart, O. Yanes, A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data, Metabolites, 2 (2012) 775-795 https://doi.org/10.3390/metabo2040775.

[13] K. Banaszewski, E. Park, I. Edirisinghe, J.C. Cappozzo, B.M. Burton-Freeman, A pilot study to investigate bioavailability of strawberry anthocyanins and characterize postprandial plasma polyphenols absorption patterns by Q-TOF LC/MS in humans, J. Berry Res., 3 (2013) 113-126 https://doi.org/10.3233/JBR-130048.

[14] S. Medina, F. Ferreres, C. Garcia-Viguera, M.-N. Horcajada, J. Orduna, M. Saviron, G. Zurek, J. Martinez-Sanz, J. Gil, A. Gil-Izquierdo, Non-targeted metabolomic approach reveals urinary metabolites linked to steroid biosynthesis pathway after ingestion of citrus juice, Food Chem., 136 (2013) 938-946 https://doi.org/10.1016/j.foodchem.2012.09.004.

[15] P. Franceschi, M. Giordan, R. Wehrens, Multiple comparisons in mass-spectrometry-based-omics technologies, TrAC, Trends Anal. Chem., 50 (2013) 11-21.

Section 5

[16] S. Lamichhane, P. Sen, A.M. Dickens, T. Hyötyläinen, M. Orešič, An overview of metabolomics data analysis: current tools and future perspectives, Comprehensive analytical chemistry, 82 (2018) 387-413 https://doi.org/10.1016/j.trac.2013.04.011.

[17] J. Godzien, M. Ciborowski, S. Angulo, C. Barbas, From numbers to a biological sense: H ow the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC-MS, Electrophoresis, 34 (2013) 2812-2826 https://doi.org/10.1002/elps.201300053.

[18] C. Manach, A. Scalbert, C. Morand, C. Rémésy, L. Jiménez, Polyphenols: food sources and bioavailability, Am. J. Clin. Nutr., 79 (2004) 727-747 https://doi.org/10.1093/ajcn/79.5.727.

[19] C. Ancillotti, L. Ciofi, D. Rossini, U. Chiuminatto, J. Stahl-Zeng, S. Orlandini, S. Furlanetto, M. Del Bubba, Liquid chromatographic/electrospray ionization quadrupole/time of flight tandem mass spectrometric study of polyphenolic composition of different Vaccinium berry species and their comparative evaluation, Anal. Bioanal. Chem., 409 (2017) 1347-1368 https://doi.org/10.1007/s00216-016-0067-y.

[20] S. Patel, Blueberry as functional food and dietary supplement: The natural way to ensure holistic health, Mediterr. J. Nutr. Metab., 7 (2014) 133-143 https://doi.org/10.3233/MNM-140013.

[21] M. Del Bubba, C. Di Serio, L. Renai, C.V.A. Scordo, L. Checchini, A. Ungar, F. Tarantini, R. Bartoletti, Vaccinium myrtillus L. extract and its native polyphenol-recombined mixture have anti-proliferative and pro-apoptotic effects on human prostate cancer cell lines, Phytother. Res., 35 (2020) 1089-1098 https://doi.org/10.1002/ptr.6879.

[22] S. Thomasset, D.P. Berry, H. Cai, K. West, T.H. Marczylo, D. Marsden, K. Brown, A. Dennison, G. Garcea, A. Miller, Pilot study of oral anthocyanins for colorectal cancer chemoprevention, Cancer Prev Res, 2 (2009) 625-633 https://doi.org/10.1158/1940-6207.CAPR-08-0201.

[23] R.P. Feliciano, G. Istas, C. Heiss, A. Rodriguez-Mateos, Plasma and urinary phenolic profiles after acute and repetitive intake of wild blueberry, Molecules, 21 (2016) 1120 https://doi.org/10.3390/molecules21091120.

[24] A. Rodriguez-Mateos, R.P. Feliciano, T. Cifuentes-Gomez, J.P. Spencer, Bioavailability of wild blueberry (poly) phenols at different levels of intake, J. Berry Res., 6 (2016) 137-148 10.3233/JBR-160123.

[25] C. Ancillotti, L. Ciofi, D. Pucci, E. Sagona, E. Giordani, S. Biricolti, M. Gori, W.A. Petrucci, F. Giardi, R. Bartoletti, Polyphenolic profiles and antioxidant and antiradical activity of Italian berries from Vaccinium myrtillus L. and Vaccinium uliginosum L. subsp. gaultherioides (Bigelow) SB Young, Food Chem., 204 (2016) 176-184 https://doi.org/10.1016/j.foodchem.2016.02.106.

[26] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol., 30 (2012) 918-920 https://doi.org/10.1038/nbt.2377.

[27] M.M. Ulaszewska, K. Trost, J. Stanstrup, K.M. Tuohy, P. Franceschi, M.F.-F. Chong, T. George, A.M. Minihane, J.A. Lovegrove, F. Mattivi, Urinary metabolomic profiling to identify biomarkers of a flavonoid-rich and flavonoid-poor fruits and vegetables diet in adults: the FLAVURS trial, Metabolomics, 12 (2016) 32 https://doi.org/10.1007/s11306-015-0935-z.

[28] M. Garcia-Aloy, M. Ulaszewska, P. Franceschi, S. Estruel-Amades, C.H. Weinert, A. Tor-Roca, M. Urpi-Sarda, F. Mattivi, C. Andres-Lacueva, Discovery of Intake Biomarkers of Lentils, Chickpeas and White Beans by Untargeted LC-MS Metabolomics in Serum and Urine, Mol. Nutr. Food Res., (2020) 1901137 https://doi.org/10.1002/mnfr.201901137.

[29] H. Zhang, M. Zhu, K.L. Ray, L. Ma, D. Zhang, Mass defect profiles of biological matrices and the general applicability of mass defect filtering for metabolite detection, Rapid Commun. Mass Spectrom., 22 (2008) 2082-2088 https://doi.org/10.1002/rcm.3585.

[30] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, Proposed minimum reporting standards for chemical analysis, Metabolomics, 3 (2007) 211-221 https://doi.org/10.1007/s11306-007-0082-2.

[31] K.M. Kim, G.N. Henderson, X. Ouyang, R.F. Frye, Y.Y. Sautin, D.I. Feig, R.J. Johnson, A sensitive and specific liquid chromatography–tandem mass spectrometry method for the determination of intracellular and extracellular uric acid, J. Chromatogr. B, 877 (2009) 2032-2038 https://doi.org/10.1016/j.jchromb.2009.05.037.

[32] C.-R. Li, X.-H. Hou, Y.-Y. Xu, W. Gao, P. Li, H. Yang, Manual annotation combined with untargeted metabolomics for chemical characterization and discrimination of two major crataegus species based on liquid chromatography quadrupole time-of-flight mass spectrometry, J. Chromatogr. A, 1612 (2020) 460628 https://doi.org/10.1016/j.chroma.2019.460628.

[33] Y. Lv, X. Liu, S. Yan, X. Liang, Y. Yang, W. Dai, W. Zhang, Metabolomic study of myocardial ischemia and intervention effects of Compound Danshen Tablets in rats using ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry, J. Pharm. Biomed. Anal., 52 (2010) 129-135 https://doi.org/10.1016/j.jpba.2009.12.016.

[34] B. Jakše, B. Jakše, M. Pajek, J. Pajek, Uric acid and plant-based nutrition, Nutrients, 11 (2019) 1736 http://dx.doi.org/10.3390/nu11081736.

[35] H. Liu, T.J. Garrett, Z. Su, C. Khoo, L. Gu, UHPLC-Q-Orbitrap-HRMS-based global metabolomics reveal metabolome modifications in plasma of young women after cranberry juice consumption, J. Nutr. Biochem., 45 (2017) 67-76 https://doi.org/10.1016/j.jnutbio.2017.03.007.

[36] N. Vázquez-Manjarrez, M. Ulaszewska, M. Garcia-Aloy, F. Mattivi, G. Praticò, L.O. Dragsted, C. Manach, Biomarkers of intake for tropical fruits, Genes Nutr, 15 (2020) 1-21.

[37] S.B. Lotito, B. Frei, Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: cause, consequence, or epiphenomenon?,

Free Radical Biol. Med., 41 (2006) 1727-1746 https://doi.org/10.1186/s12263-020-00670-4.

[38] M. Godycki-Cwirko, M. Krol, B. Krol, A. Zwolinska, K. Kolodziejczyk, M. Kasielski, G. Padula, J. Grębocki, P. Kazimierska, M. Miatkowski, Uric acid but not apple polyphenols is responsible for the rise of plasma antioxidant activity after apple juice consumption in healthy subjects, J. Am. Coll. Nutr., 29 (2010) 397-406 https://doi.org/10.1080/07315724.2010.10719857.

[39] S. Martínez-López, B. Sarriá, M. Gómez-Juaristi, L. Goya, R. Mateos, L. Bravo-Clemente, Theobromine, caffeine, and theophylline metabolites in human plasma and urine after consumption of soluble cocoa products with different methylxanthine contents, Food Res. Int., 63 (2014) 446-455 https://doi.org/10.1016/j.foodres.2014.03.009.

[40] J. Oracz, E. Nebesny, D. Zyzelewicz, G. Budryn, B. Luzak, Bioavailability and metabolism of selected cocoa bioactive compounds: A comprehensive review, Crit. Rev. Food Sci. Nutr., 60 (2020) 1947-1985 https://doi.org/10.1080/10408398.2019.1619160.

[41] R.C. Pimpao, M.R. Ventura, R.B. Ferreira, G. Williamson, C.N. Santos, Phenolic sulfates as new and highly abundant metabolites in human plasma after ingestion of a mixed berry fruit puree, Br. J. Nutr., 113 (2015) 454-463 https://doi.org/10.1017/S0007114514003511.

[42] L. Bresciani, D. Martini, P. Mena, M. Tassotti, L. Calani, G. Brigati, F. Brighenti, S. Holasek, D.-E. Malliga, M. Lamprecht, Absorption profile of (poly) phenolic compounds after consumption of three food supplements containing 36 different fruits, vegetables, and berries, Nutrients, 9 (2017) 194 https://doi.org/10.3390/nu9030194.

[43] I.B. Jaganath, I.B. Jaganath, W. Mullen, C.A. Edwards, A. Crozier, The relative contribution of the small and large intestine to the absorption and metabolism of rutin in man, Free Radical Res., 40 (2006) 1035-1046 https://doi.org/10.1080/10715760600771400.

[44] T. Nurmi, J. Mursu, M. Heinonen, A. Nurmi, R. Hiltunen, S. Voutilainen, Metabolism of berry anthocyanins to phenolic acids in humans, J. Agric. Food. Chem., 57 (2009) 2274-2281 https://doi.org/10.1021/jf8035116.

[45] Y. Hövelmann, L. Lewin, K. Steinert, F. Hübner, H.U. Humpf, Mass Spectrometry-Based Analysis of Urinary Biomarkers for Dietary Tomato Intake, Mol. Nutr. Food Res., 64 (2020) 2000011 https://doi.org/10.1002/mnfr.202000011.

[46] T. Herraiz, Analysis of the bioactive alkaloids tetrahydro-β-carboline and β-carboline in food, J. Chromatogr. A, 881 (2000) 483-499 https://doi.org/10.1016/S0021-9673(99)01313-8.

[47] I.I. Koleva, T.A. van Beek, A.E. Soffers, B. Dusemund, I.M. Rietjens, Alkaloids in the human food chain–natural occurrence and possible adverse effects, Mol. Nutr. Food Res., 56 (2012) 30-52 https://doi.org/10.1002/mnfr.201100165.

[48] A. Akagić, A.V. Oras, S. Oručević Žuljević, N. Spaho, P. Drkenda, A. Bijedić, S. Memić, M. Hudina, Geographic Variability of Sugars and Organic

Acids in Selected Wild Fruit Species, Foods, 9 (2020) 462 https://doi.org/10.3390/foods9040462.

[49] M. Mikulic-Petkovsek, V. Schmitzer, A. Slatnar, F. Stampar, R. Veberic, Composition of sugars, organic acids, and total phenolics in 25 wild or cultivated berry species, J. Food Sci., 77 (2012) C1064-C1070 https://doi.org/10.1111/j.1750-3841.2012.02896.x.

[50] M.M. Phillips, R.J. Case, C.A. Rimmer, L.C. Sander, K.E. Sharpless, S.A. Wise, J.H. Yen, Determination of organic acids in Vaccinium berry standard reference materials, Anal. Bioanal. Chem., 398 (2010) 425-434 https://doi.org/10.1007/s00216-010-3916-0.

[51] K. Karppinen, E. Hirvelä, T. Nevala, N. Sipari, M. Suokas, L. Jaakola, Changes in the abscisic acid levels and related gene expression during fruit development and ripening in bilberry (Vaccinium myrtillus L.), Phytochemistry, 95 (2013) 127-134 https://doi.org/10.1016/j.phytochem.2013.06.023.

[52] S. Medjakovic, A. Jungbauer, Pomegranate: a fruit that ameliorates metabolic syndrome, Food Funct, 4 (2013) 19-39 https://doi.org/10.1039/C2FO30034F.

[53] H. Zhang, Z.F. Ma, X. Luo, X. Li, Effects of mulberry fruit (Morus alba L.) consumption on health outcomes: A mini-review, Antioxidants, 7 (2018) 69 https://doi.org/10.3390/antiox7050069.

6 Combining Feature-Based Molecular Networking and Contextual Mass Spectral Libraries to Decipher Nutrimetabolomics Profiles

Supplementary Materials:

https://www.mdpi.com/article/10.3390/metabo12101005/s1

Abstract

Untargeted metabolomics approaches deal with complex data hindering structural information for the comprehensive analysis of unknown metabolite features. We investigated the metabolite discovery capacity and the possible extension of the annotation coverage of the Feature-Based Molecular Networking (FBMN) approach by adding two novel nutritionally-relevant (contextual) mass spectral libraries to the existing public ones, as compared to widely-used open-source annotation protocols. Two contextual mass spectral libraries in positive and negative ionization mode of ~300 reference molecules relevant for plant-based nutrikinetic studies were created and made publicly available through the GNPS platform. The postprandial urinary metabolome analysis within the intervention of Vaccinium supplements was selected as a case study. Following the FBMN approach in combination with the added contextual mass spectral libraries, 67 berry-related and human endogenous metabolites were annotated, achieving a structural annotation coverage comparable to or higher than existing non-commercial annotation workflows. To further exploit the quantitative data obtained within the FBMN environment, the postprandial behavior of the annotated metabolites was analyzed with Pearson product-moment correlation. This simple chemometric tool linked several molecular families with phase II and phase I metabolism. The proposed approach is a powerful strategy to employ in longitudinal studies since it reduces the unknown chemical space by boosting the annotation power to characterize biochemically relevant metabolites in human biofluids.

Keywords: human urine, liquid chromatography, untargeted mass spectrometry, computational metabolomics, chemometrics, bioinformatics

6.1   Introduction

Untargeted tandem mass spectrometry (MS/MS) is one of the most widely used analytical techniques in metabolomics, allowing for the generation of information-rich mass spectral datasets and the identification of metabolic

biomarkers in biological complex mixtures [1,2], also thanks to the coupling with separation techniques such as liquid chromatography (LC). Despite the wide application of hyphenated LC-MS/MS platforms, the annotation of biologically relevant metabolites (i.e., biomarkers) is strongly hampered by the complexity of the metabolome and metabolomics data processing and annotation [3]. The annotation process is a pivotal step in untargeted metabolomics that often represents a bottleneck in the process of obtaining biological information and discovering biomarkers. To streamline the metabolite annotation process, metabolomics guidelines have been proposed for the accurate identification and assignment of a metabolite feature [4,5], i.e., through peak picking, mass spectral deconvolution, determination of molecular ions by adduct detection, and fragmentation pattern (MS/MS) analysis [6]. Despite these efforts, the risk of missing relevant information and drawing incorrect conclusions remains relatively high, due to incorrect MS and MS/MS interpretations when matching experimental spectra with available spectral libraries. To aid in structural interpretation, the identification of MS/MS spectral similarities within a given dataset can support the discovery of structurally related metabolites, which plausibly share the same metabolic pathway and/or substructure [7], thus strengthening the biological meaning of the annotations. In this context, molecular networking (MN) has gained large attention, thanks to the efficient and rapid identification of several molecular families within complex mixtures, providing a visual overview of all the precursor ions, grouped according to their structural relationships, as deduced by their mass fragmentation spectra during an MS/MS experiment [8]. MN uses an unsupervised vector-based computational algorithm to organize molecular ions (i.e., clusters or nodes) into a network of molecular families that share spectral similarities among their MS/MS spectra. At the same time, structural annotation is performed through the Global Natural Products Social Molecular Networking (GNPS) bioinformatics platform [8], which is linked to many mass spectral libraries available as public repository of mass spectra and metadata (i.e., GNPS-MassIVE). Considering the recent growth of

public mass spectral libraries, it is expected an increase of the annotation capability (level II or level III) of biologically relevant molecules in comparison with traditional biomarker discovery workflows [9]. MN has been applied in several untargeted LC-MS/MS studies, mainly focusing on phytochemical composition analysis [8], and less frequently on drug metabolism [10], and nutrimetabolomics [11] in human biofluids. Recently, MN has been extended by its combination with standard feature detection tools into the Feature-Based Molecular Networking (FBMN) workflow that is capable to resolve isomers and incorporate quantitative information (e.g., spectral counts, chromatographic peak areas, etc.), increasing the link between peak picking algorithms and in silico annotation tools [12]. Until now, FBMN has been successfully applied in various fields of metabolomics, allowing level II/level III identification of transformation products of organic micropollutants in water samples [13], native plant constituents [14,15,16], and endogenous urinary metabolites [17]. However, mass spectral library matching is generally performed by the comparison with mass spectral libraries containing MS/MS spectra acquired under a wide range of instrumental conditions (e.g., time-of-flight, orbitrap, hybrid ion traps, etc.) and collision energies used, with different curation protocols providing different mass accuracy levels [13,16], thus suffering from limited reliability of the annotation due to differences in observed mass fragments and their intensity ratios. This issue can be managed by implementing better contextualized libraries containing reference spectra of study-related compounds and acquired under experimental conditions equal to or comparable to the experimental data being analyzed. Finally, FBMN has the hitherto unexploited potential in biomarker research to provide quantitative data of the structurally annotated (and unannotated) features, thus complementing the traditional biomarker discovery procedure with a chemometric protocol that allows establishing their biological significance.

This research investigates the discovery capacity and the extension of the annotation coverage of the FBMN approach, in comparison with a commonly adopted manual annotation of selected significant m/z features [18]. To this end,

the FBMN workflow was applied to deconvoluted and aligned high-resolution LC-MS/MS files of postprandial urine samples from a two-arms intervention study on the intake of Vaccinium myrtillus (VM) and Vaccinium corymbosum (VC) berry supplements. As far as we are aware, this represents the first nutrimetabolomics application of FBMN to the identification of postprandial endogenous and exogenous metabolites. The MS/MS spectra that were acquired in both negative ionization (NI) and positive ionization (PI) data dependent acquisition modes were compared with the available GNPS libraries. An extensive comparative analysis was done to compare various FBMN parameter settings to arrive at optimal settings for structural annotation purposes in the nutrimetabolomics setting. Furthermore, to support automated nutrimetabolomics annotation workflows, two novel NI and PI contextualized "Nutri-Metabolomics" mass spectral libraries were constructed and made available uniquely on GNPS, each containing MS/MS spectra of about three-hundred food-related human metabolites, acquired under the same mass spectrometric conditions as the study samples. These mass spectral libraries are a fruit of several years of investigations on human responses to dietary interventions at the Edmund Mach Foundation (Italy), and include phase I and phase II human metabolites, as well as food constituents. Special attention was given to microbial metabolites resulting from mixed human and microbiome interaction such as small phenolic acids, phenylacetic acids, phenylpropionic acids, indoles, and carbolines, as well as bile acids. Other classes include sulfate and glucuronides conjugates of common food constituents such as caffeic acid glucuronide, dihydroferulic acid sulfate, isoferulic glucuronide, etc. Several aroma compounds were included to facilitate substructure matching, as those were observed in biological fluids in conjugated form (monoterpenoids, safranal, furfuran, fenchyl alcohol etc.). Finally, the spectral library offers specific advanced glycation end-products including pyrraline, furosine and more.

In the current study, the mass spectral library creation aimed at increasing (i) the accuracy in the annotation thanks to a better match of instrumental metadata such

as detector and collision energy, (ii) nutrimetabolomics knowledge on postprandial analysis of biological samples and plant-based food intake. Additionally, the quantitative data within NI and PI FBMN networks were exploited to gain insights into (i) metabolites characterized by different postprandial kinetics and (ii) the relative dietary contribution of VM and VC interventions of the identified metabolites.

6.2   Materials and Methods

6.2.1   Chemicals and reagents

Full purchase details of solvents and standards used are reported in section S1 of the *Supplementary materials*. The complete list of the reference standard adopted to build the "Nutri-Metabolomics" libraries, in both NI and PI modes, is shown in the *List of Reference Standards used to build the libraries.xlsx* file in the *Supplementary materials*.

6.2.2   Study design, sample extraction, and LC-MS/MS analysis

The datasets analyzed in this research are part of a more comprehensive clinical intervention trial, based on the hematic and urine biomarker discovery on the intake of VM and VC [18,19]. Urine samples of each volunteer ($n = 10$ for each intervention) were collected at baseline and 30, 60, 120, 240, and 360 min after VM or VC supplement intake. Pooled urine samples were also collected 24 h and 48 h after supplement intake. Details of supplements characterization (**Table S1**), as well as study design are reported in Section S2 of the *Supplementary Materials*. Urine samples were extracted and analyzed as reported elsewhere [19]. The entire procedures of extraction and LC-MS/MS analysis of urine samples are reported in Sections S3 and S4 of the *Supplementary Materials*, respectively. The entire sample set was acquired in full scan mode, collecting high quality data for an appropriate statistical analysis, as well as in data dependent acquisition (DDA) mode, to leverage large quantities of MS/MS data for structural investigation preserving the kinetic heritage of the study design.

6.2.3   Data pre-processing

The full scan files were processed and analyzed as previously reported [18]. Additionally, the data-dependent spectra files (including blanks) were converted from the .raw to .mzML MS convert by ProteoWizard (https://proteowizard.sourceforge.io) (accessed on 15 April 2021). Further data processing was performed with MZmine 2 software [20], separately for NI and PI datasets. Data pre-processing included the following steps: mass detection, chromatogram reconstruction and deconvolution, isotope grouping, alignment and gap filling. Subsequently, the aligned feature lists were exported as MS/MS files (.mgf format) and quantification tables (.csv format of aligned features and related chromatographic peak areas), according to GNPS documentation on FBMN (https://ccms-ucsd.github.io/GNPSDocumentation/) (accessed on 21 April 2021).

6.2.4   Data availability: MassIVE repository, metadata and GNPS jobs

Data in .mzML format are available on-line on GNPS infrastructure (MSV000088336). The metadata describing file/sample properties were entered manually for all samples and organized in two different files according to the acquisition polarity of the uploaded MassIVE datasets, following the GNPS guidelines   (https://ccms-ucsd.github.io/GNPSDocumentation/metadata/) (accessed on 21 April 2021). In detail, metadata consisted of three descriptive categories, (i) spectrum file name (the same of acquired raw data), (ii) type of supplement (VM or VC), and (iii) related time point after intake. These elements are required to get a correct grouping within FBMN for quantitative analysis (see the Metadata and Library Information.xlsx file in the *Supplementary Materials*). For the upload on GNPS, metadata files were converted to .tsv files. The FBMN analysis are available at the following links:

PI:
https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a981ebd40809453ebe1524ff1fc8e265
(accessed on 27 June 2021).

NI:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0a239e71bb2045c292c4c96f4501249c
(accessed on 27 June 2021).

6.2.5   *"Nutri-Metabolomics" library building and implementation*

The analytical standards used to build the in-house libraries were acquired in the same MS/MS conditions as study samples (replicated three times), which are reported in Section S4 of the *Supplementary Materials*. GNPS provides a platform to build MS/MS spectral libraries, requiring good quality MS/MS spectra and annotation spread sheets containing key and machine-readable descriptors such as file name, compound name, SMILES, InChiKey, PubMed. To build the library, only pure analytical standards were used, thus no putative or un-known compounds are present in the files. Two annotation spread sheets were built in NI and PI, containing 319 and 339 injected compounds, respectively (see the Metadata and Library Information.xlsx file in the *Supplementary Materials*). Analysis of standards included their separation on the chromatographic column; however, a retention time match is not supported in GNPS and therefore this information was used manually when needed. The "Batch Validator Workflow" [21] step was run to evaluate the correct match between spreadsheets (dropped as .csv files), and original spectra. The completed libraries can be found in the public spectral library collection of GNPS named as "Nutri-Metabolomics".

6.2.6   *Molecular networking analyses*

Molecular networks were obtained following the online workflow on the GNPS web-platform (https://gnps.ucsd.edu/) (accessed on 21 April 2021). FBMN was performed adopting the most suitable basic and advanced networking options, selected through the recommended network qualitative optimization by classical MN (see Section 3.1), for NI and PI dataset exported from MZmine 2 software. The detailed investigation of MN options is reported in Section S5 of the Supplementary Materials. The most appropriate input parameters were set as follows: NI were analyzed using precursor ion mass tolerance (PIMT) and

fragment ion mass tolerance (FIMT) equal to 0.1 Da and 0.01 Da, respectively. The other parameters were set as follows: minimum matched fragment ions = 3, networking cosine score > 0.6, library cosine score > 0.5, and minimum library shared peaks = 3. PI dataset was processed adopting PIMT = 0.05, FIMT = 0.05, minimum matched fragment ions = 3, networking cosine score > 0.5, library cosine score > 0.3, and minimum library shared peaks = 3. Network analysis and quantitative results were investigated and exported adopting Cytoscape environment [22]. Moreover, unknown nodes were annotated with putative molecular structures by manual annotation based on: (i) mass difference between identified and unknown node, (ii) precursor ion mass accuracy, and (iii) fragmentation patterns in MS/MS spectra (see Section S6 of *Supplementary Materials*).

6.2.7 *Analysis of postprandial kinetics*

Reinjection of the entire dataset in DDA fashion enabled the exploitation of postprandial kinetics data. To extract the postprandial information from FBMN, the Pearson product-moment correlation (PPMC) analysis was performed, using the "corrplot: A visualization of a correlation matrix" package implemented in R (https://cran.r-project.org/) (accessed on 28 July 2021), thus estimating the linear correlation between the mean chromatographic peak area of identified nodes and time points. The quantitative FBMN data used for the correlation analysis were extracted from the "node table" of the Cytoscape environment, built using the loaded metadata for both NI and PI datasets. Statistically significant (p-value ≤ 0.05) PPMC coefficients (r) were used to discriminate early (1–2 h postprandial) from late (approximately 4 h and more postprandial) occurring postprandial metabolites, which are commonly considered as the result of phase II or phase I metabolism, respectively [23]. Accordingly, positive and negative r-values indicated nodes associated to phase I (late postprandial) and phase II (early postprandial) metabolism, respectively. A limitation of using PPMC within FBMN was the absence of sample normalization as this functionality is currently not available. Findings from this step were compared to those obtained through

the PPMC analysis of longitudinal variations of the chromatographic area of aligned features (i.e., outside FBMN), as a control strategy. It should be highlighted that, although the PPMC coefficients can be associated with the metabolism phase, its relation to the specific food intake remains elusive without further biochemical interpretations. Simultaneously, full scan data underwent the conventional data processing, as previously described [17]. Briefly, biomarkers of food intake in postprandial responses were selected by applying selected R packages to full scan data [24], according to the following two-step procedure: (i) verification of increasing trend along time points and (ii) calculation of AUC curves and intra-intervention discrimination. Statistically significant features were annotated manually with use of on-line spectra databases such as mzCloud and HMDB. Details of this procedure are reported in Section S5 of *Supplementary Materials*.

6.3   Results and Discussion

The NI and PI datasets were treated following the workflow illustrated in **Figure 1**, which integrates the PPMC analysis of postprandial kinetics within the FBMN environment. However, since FBMN extracts only the mean values of the chromatographic area as quantitative data for PPMC analysis, thereby losing knowledge of inter-individual variability, the variance of metabolite feature abundance among volunteers was investigated at each time point as a control, before applying the FBMN workflow. Accordingly, the coefficient of variation (CV%) of chromatographic areas of each aligned feature within a same time point was calculated, highlighting a strong variability (CV% approximately in the range of 30-300% and median higher than 100% in most cases). These findings highlighted the importance of evaluating the results of PPMC analysis at the population level.
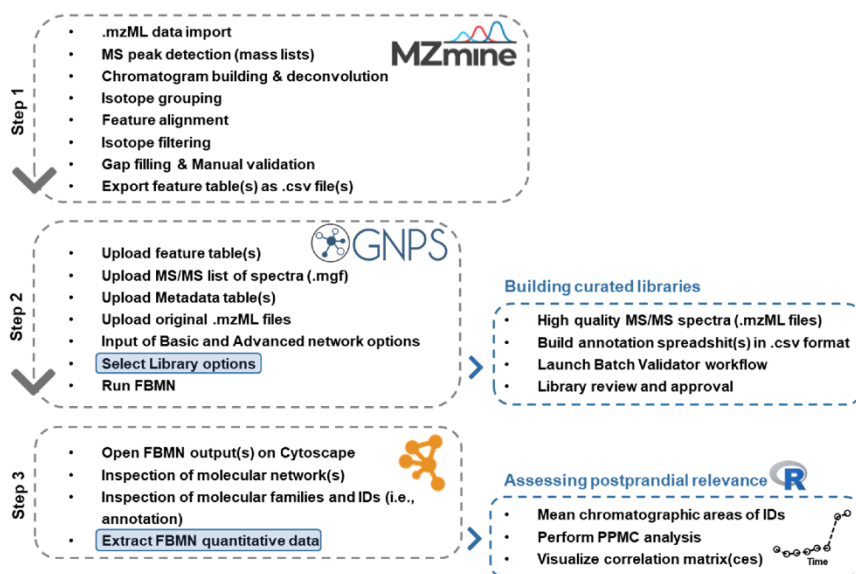
Figure 1 – Schematic representation of the available (grey dashed lines) and proposed (blue dashed lines) workflows of data management, applied by the combination of Feature-Based Molecular Networking and the "Nutri-Metabolomics" mass spectral libraries that were manually curated.

6.3.1   *Optimization of the input parameters for network analysis*

Before running FBMN, various networking basic and advanced options must be investigated to find out the most suitable parameters to perform the MN analysis. To properly evaluate the effect of input parameters, the total number of nodes (precursor ions with identical fragmentation pattern, i.e., consensus spectrum), edges (i.e., node connections related to structural similarities), identified compounds (IDs, i.e., annotated through spectral library matching), and spectral families (i.e., the groups or clusters, also referred as molecular families), were analyzed in both NI and PI datasets and the results are reported in **Figure S1** of the *Supplementary Materials*. In this regard, increasing PIMT value, the number of nodes, edges, and spectral families decreased, whereas the number of IDs showed a predominantly increasing trend, mainly due to the less strict conditions as consensus spectra got merged (i.e., considering different isobaric compounds as one) at increasing PIMT. Hence, to keep a reliable number of nodes and spectral families without significantly affecting the number of IDs, PIMT was set at 0.1

Da and 0.05 Da for NI and PI, respectively. FIMT exerted an effect on the output variables like that of PIMT, except for the total number of edges, which increased by increasing values of FIMT. Due to the loss of accuracy in node networking for high FIMT, values of 0.01 and 0.05 were chosen for NI and PI datasets, respectively. The number of minimum matched fragment ions was set at 3 for both NI and PI for two reasons: (i) Its increase exerts a significant reduction of the number of nodes with an ID and their reliability, (ii) many food-derived metabolites have only a few characteristic mass fragments. Cosine scores for networking and library matching affected mainly the number of spectral families and of IDs, respectively. A good compromise between these two outputs was obtained by setting the networking and library matching cosine score thresholds at 0.6 for NI and 0.5 for PI. Finally, the number of minimum library shared peaks was set at 3, because higher values of this parameter were responsible for a drastic reduction of IDs, similarly to what was observed for the number of matched fragments.

6.3.2   *FBMN annotation of NI and PI datasets*

FBMN workflow applied to NI and PI datasets combined with aligned feature lists and quantitative tables exported from data pre-processing, was able to remove the 57% and 27% of NI and PI redundant IDs (i.e., artefacts like duplicated features) found by classical MN, respectively. As first result, the effect of including context specific "Nutri-Metabolomics" mass spectral libraries in the annotation workflow was evaluated by applying the FBMN protocol in their presence and absence (i.e., GNPS libraries "only"). Indeed, substantial advantages were observed upon using the dedicated mass spectral libraries, i.e., the increase of (i) 20%, 48%, in the number of IDs (**Figure S2A**) and (ii) 62.5%, 34%, in the number of IDs with a mass error < 5 ppm (**Figure S2B**), for NI and PI datasets, respectively. Additionally, the use of the "Nutri-Metabolomics" libraries solved two mis-annotations (i.e., incorrect annotation of nodes) in the NI datasets. These results highlight the importance of applying the FBMN annotation strategy in combination with contextual libraries, i.e., containing true reference standards that

are relevant for the application of interest and analyzed under the same instrumental conditions adopted for the analysis of real samples. The FBMN network of the NI dataset consisted of 545 nodes and 799 edges, with a total number of connected components equal to 307, corresponding to 65 spectral families, whereas molecular networking of the PI dataset resulted in 5079 nodes and 6904 edges, with a total number of connected components equal to 3543 (i.e., 663 spectral families). The ID lists obtained from the library matching in both NI and PI datasets contained 39 and 384 unique annotated compounds, respectively, which were checked for mass accuracy to be around or lower than 5 ppm. **Table S2** (see Section S5 of the *Supplementary Materials*) reports the metabolites identified by library matching (based on cosine score similarity) of nodes within and outside molecular families (the latter are typically called singletons) from both NI (24 IDs) and PI (43 IDs) datasets, characterized by the lowest mass error (Δ ppm). Even though the FBMN approach has specific methodological inputs and results that differentiate it from commonly used workflows in untargeted nutrimetabolomics, it is interesting to compare the discovery capacity and annotation coverage obtained with other approaches. For this purpose, the FBMN was compared against two widely used annotation protocols: (i) MZmine Library Search and (ii) statistical-based feature selection followed by manual annotation (see Section S5 of the *Supplementary Materials*) [18]. It should be emphasized that the compared workflows differ substantially as per their rationale. MZmine Library Search workflow matches each row of the NI and PI feature lists (used also for FBMN) against the imported spectral library. To make a consistent comparison with the annotation performed with FBMN, the "ALL-GNPS" library was used. The conventional protocol aims at selecting only statistically significant m/z features from full scan data, followed by manual annotation using the MS/MS spectra often obtained in targeted mode. In contrast with the presented approaches, FBMN explores all available MS/MS data from the DDA metabolomics profiles (taking advantage of all structural annotations that can be made), annotating them against mass spectral libraries. Only then, further statistical analysis is performed

to discover their potential postprandial relevance. Thus, the direct comparison of these annotation and prioritization workflows is not and will never be straightforward; yet, here we highlight some relevant aspects. **Table 1** shows the final number of IDs found adopting the three approaches. MZmine Library Search workflow provided the metabolite annotation with 26 and 49 unique IDs in NI and PI datasets, respectively, with cosine similarity scores (isotopic pattern at full scan level) higher than 0.7. The number of IDs identified by this approach was comparable with the results of the applied FBMN workflow, and several metabolite categories were commonly annotated by the two procedures (data not shown), such as hippuric acids, catechols, and derivatives of phenylacetic acid, coumaric acid, indoles, and hydroxybenzoic acid. However, due to the format of our data unsuitable for MS/MS-based mass spectral matching within MzMine (i.e., incomplete mass lists for the MS/MS scans), the MZmine-based approach relied on precursor m/z and isotope pattern matching, thus possibly resulting in a less reliable annotations due to the limited structural information. The statistical-based/manual annotation method resulted in 50 and 106 statistically significant m/z features in PI and NI datasets, respectively, corresponding to 24 metabolite features after manual checking. Manual structure elucidation putatively identified 18 metabolites (12 in NI and 6 in PI datasets), while 6 metabolites remained unknown (see **Table S3**). Using FBMN, a higher number of metabolites was putatively annotated, i.e., 24 IDs in NI, and 43 IDs in PI, when compared to the statistical-based/manual annotation approach. These differences were due to both (i) the automatic query (intrinsic of FBMN) of all publicly available mass spectral libraries, including "Nutri-Metabolomics" ones, and (ii) the different strategies to select the metabolite features to be annotated. In fact, the conventional approach processes the NI and PI datasets to highlight physiologically relevant features, before their annotation is performed by unqueried matching with analytical standards available in on-line spectral libraries. On the contrary, FBMN automatically generates a list of IDs, which is then refined by applying, for example, a mass accuracy threshold, in combination with the use of mass spectral

similarity scoring (i.e., modified cosine score), as presented in this study. Despite these methodological differences, hydroxyhippuric acid and dihydrocaffeic acid glucuronide were identified with both approaches. Moreover, the conventional postprandial analysis confirmed the FBMN identification of structurally-related metabolites significantly altered upon berry intake, belonging to furoic and abscisic acid derivatives, hydroxy and/or methoxy benzoic acids. By contrast with FBMN, the conventional protocol for postprandial analysis identified the metabolite categories of valerolactone and valeric acid derivatives (see Section S5 and Figure S3 of the Supplementary Materials), which are well-known colon-derived catabolites of flavanols [25]. These metabolite features were found also inside the FBMN molecular networks; however, they were not structurally characterized as such, due to their absence in the "Nutri-Metabolomics" and other mass spectral libraries. These findings highlighted the importance of expanding the coverage of online spectral repositories to boost metabolite annotations.

Table 1 – Number of IDs annotated by Feature-Based Molecular Networking (FBMN) of NI and PI datasets, including the developed "Nutri-Metabolomics" mass spectral libraries, in comparison with the annotation performed with (i) MZmine Library Search using GNPS compatible mass spectral libraries (ALL_GNPS, https://gnps-external.ucsd.edu/gnpslibrary) (accessed on 5 September 2022) and with (ii) the statistical-based approach followed by manual annotation, reported in Section S5 of the *Supplementary materials*.

Number of IDs	NI	PI
MZmine[1]	26	49
Statistical-based approach & manual annotation	12	6
FBMN	24	43

[1] Library search performed at full scan MS level using *m/z* and isotope pattern matching.

6.3.3   *VM and VC relative contributions to the postprandial metabolome*

Categorization of NI and PI metadata based on VM and VC interventions (see Section 2.4 for details) allowed for the separate storage of spectral counts (i.e., the number of mass spectra recorded for a node) of each ID precursor ion. This information was used here for assessing the VM and VC relative contributions of each ID to the postprandial metabolome, by the representation of a pie chart (see **Figure 3** and **Figure 4** in Section 3.5). Moreover, a preliminary and descriptive

contribution to the annotated urinary metabolome of VM and VC interventions can be estimated. Interestingly, VM and VC interventions exhibited an opposite feature occurrence in the two ionization datasets, highlighting the importance of investigating both polarity modes. In detail, NI IDs resulted in a higher postprandial occurrence after the intake of VC supplement ($62 \pm 6\%$ vs. $38 \pm 4\%$ for VC and VM), whereas for PI dataset, a slight predominance was found for VM ($54 \pm 2\%$ vs. $46 \pm 2\%$ for VM and VC).

6.3.4   PPMC analysis of postprandial kinetics at the population level

Longitudinal data analyzed by FBMN approach allows for additional data exploration to highlight the specificity of food intake as well as the "background" metabolism, since no feature selection is performed. Accordingly, the PPMC analysis was performed on the mean values of chromatographic area of each ID as a function of time. Following this analysis, 65.7% of the annotated metabolites (i.e., 44 IDs on a total of 67) showed a statistically significant trend approximating an increasing or decreasing postprandial response, thus highlighting the reliability of this approach. Among the significant correlated metabolites, 35 IDs showed a positive coefficient (r-values) and were therefore associated to phase I metabolism, whilst 9 IDs were characterized by negative r-values, suggesting a phase II metabolism. **Figure 2A,B** illustrates two representative postprandial trends of IDs corresponding to significant negative and positive r-values, respectively The remaining 23 metabolites exhibited a non-linear and not significant trend, as shown in the two representative examples of **Figure 2C,D**. The postprandial behavior of these IDs cannot therefore be assigned through this approach and requires a qualitative investigation (plots of chromatographic areas vs timepoints) and/or a dedicated treatment outside the FBMN environment. As stated above, these results were based on the correlation analysis of mean values of chromatographic areas, i.e., without considering the dispersion of individual data around the mean. To evaluate the impact of the extent of this dispersion on the statistical significance of linear correlations, the data obtained for each volunteer and for each annotated feature were submitted to PPMC analysis outside

the FBMN environment (i.e., using the data from MZmine feature lists). For 25 IDs out of the 44 IDs found to be significant based on mean chromatographic areas, the statistical significance of the r-values was confirmed, notwithstanding the high variability observed in the peak area datasets. These results encourage the applicability of the postprandial analysis proposed here at least as a first immediate screening of the postprandial behaviour of annotated metabolites, capturing their metabolic trends over time. It is of note that this approach would produce more accurate results when a lower dispersion of individual data around the average value is observed; and to achieve this, increasing the sample size may be of help.

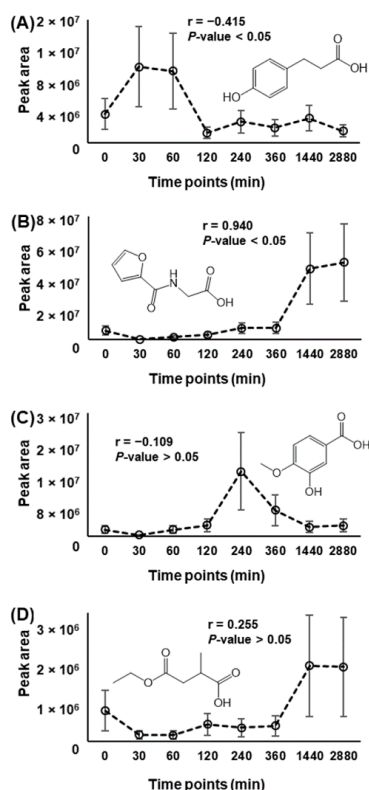**Figure 2** – Representative examples of the use of PPMC coefficients (r) to evaluate phase I and II metabolisms of annotated metabolites (Table S2), as described in paragraph 2.7 of the main text. (A) Significant and negative r-value indicating a phase II metabolism. (B) Significant and positive r-value indicating a phase I metabolism. (C) Non-significant and negative r-value. (D) Non-significant and positive r-value.

6.3.5   Nutrimetabolomics outcomes from FBMN molecular networks

Figure S4 of the *Supplementary Materials* shows representative examples of the structural modification involved in phase I and II metabolism of well-known VM and VC native constituents [26], in association with the annotated metabolites and their significant PPMC r-values. Accordingly, potential metabolic modifications such as conjugations (e.g., glucuronidation) and additions (e.g., methylation) should be expected to undergo in-source hydrolysis and dissociation, leading to accurate annotations, but losing a relevant structural information. To limit these drawbacks, a robust network inspection was performed to ensure a reliable annotation. Within NI dataset (24 IDs, see **Table S2**), four singletons were identified through spectral matching with a good mass accuracy: azelaic acid, galacturonic acid, glutamine, and ethoxy-oxobutenoic acid. The occurrence of galacturonic acid and glutamine can be addressed to in-source dissociation of glycosidic and peptidic bonds of metabolite conjugations. **Figure 3** illustrates the molecular families in which at least one of the remaining 20 metabolites was annotated. These metabolites were grouped according to their postprandial kinetics, as assessed by statistically significant r-values. In **Figure 3**, the structure of unknown nodes labelled with a gear was proposed as level III identification by the analysis of their MS/MS spectra the hypothesized scheme of fragmentation (Figure S5 of the *Supplementary Materials*). About the 50% of the identified structures was characterized by molecular scaffolds related to cinnamic and dihydrocinnamic acids. Interestingly, among unknown nodes, a relevant number of putative glucuronide derivatives was easily recognized by the occurrence in the MS/MS spectra of peaks at *m/z* 175.02 and 113.02, typical of glucuronic acid (**Figure S5**). Two nodes highlighted in one box in **Figure 3** were recognized as a molecular family related to abscisic acid glucuronide derivatives. The ID occurring in this family, was at first addressed as dihydroxy-diphenylphenoxy-trihydroxyoxane-carboxylic acid, with a mass error of about 128 ppm. However, the inspection of its MS/MS spectra (see **Figure S6** of the *Supplementary*

Materials) led to a more accurate putative annotation of this node as methoxyabscisic acid glucuronide ($\Delta = 2.6$ ppm).
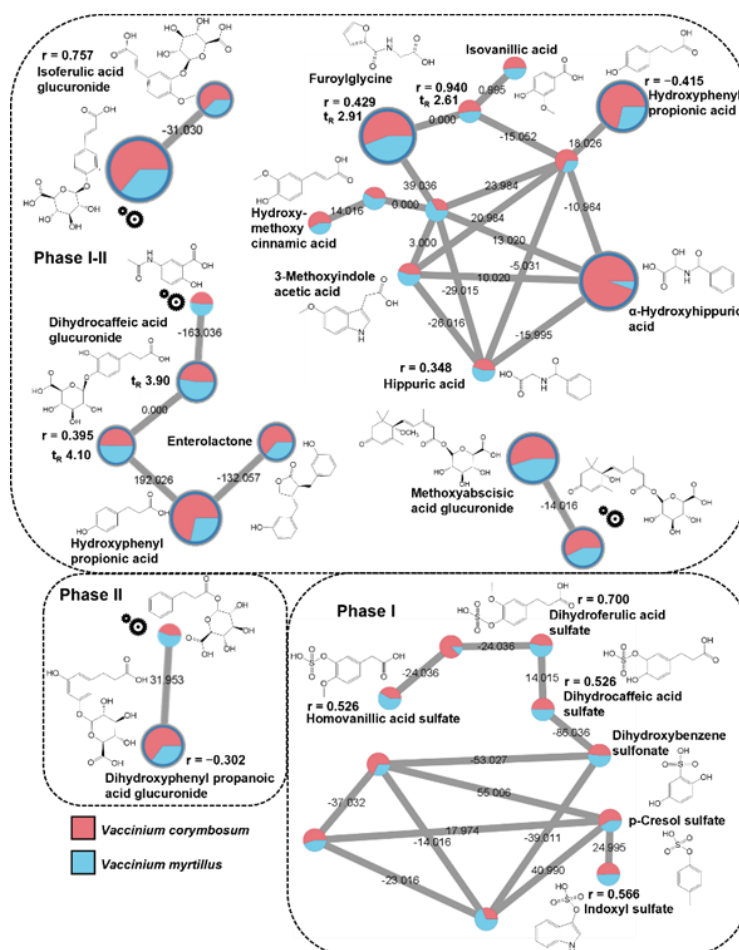
Figure 3 – Extracted networks of identified metabolites in negative ionisation mode listed in Table S2, belonging to the category of (poly)phenolic compounds, abscisic acid, and their glucuronide or sulfate derivatives. Dashed boxes group the identified metabolites according to phase I and II metabolism following PPMC analysis. The "gear" symbols refer to the putative structure identi-fied by manual investigation as reported in paragraph 2.6 of the main text. Statistically significant Pearson correlation coefficients (r) are reported. Edge labels refer to the mass difference between two nodes.

In addition, the hypothesized structure of the linked node was consistent with abscisic acid glucuronide (**Figure S5**), which was already putatively identified in a previous study by a conventional annotation workflow [18]. Among the

identified molecular families in the NI dataset, some of them exhibited a mixed metabolic contribution (i.e., phase I–II). In detail, isoferulic acid glucuronide showed a positive and significant PPMC correlation (r = 0.757), but was linked with a node exhibiting an opposite postprandial behavior (peak area vs. time points, data not shown), thus suggesting a phase I-II mixed contribution. An analogous mixed metabolic contribution can be also proposed for the abscisic acid spectral family since the methoxyabscisic acid glucuronide is most likely associated to phase I due to the methylation of the hydroxyl group, whereas the node putatively associated to the glucuronide derivative of abscisic acid, is related to phase II. The molecular family containing hydroxyphenyl propionic and hydroxy-methoxy cinnamic acids was characterized by peculiar structural relationships and depicted a heterogeneous metabolic contribution. In fact, the postprandial analysis of the identified nodes evidenced a phase I expression (0.348 < r < 0.940) for most metabolites [19,27], with the only exception of hydroxyphenyl propionic acid, for which a phase II metabolism can be suggested, based on its r-value (−0.415). Even though this metabolic association to phase II metabolism is apparently questionable due the lack of a conjugated group, the analysis of the full scan spectra evidenced an in-source fragmentation of the sulfate derivative of the hydroxyphenyl propionic acid (*m/z* = 263.02 Da), thus confirming the phase II metabolic attribution. A detailed analysis of this molecular family evidenced also that the compound annotated as hydroxy-methoxycinnamic acid probably underwent in-source dissociation, since in the same t_R range, an ion at *m/z* 273 fragmented in *m/z* 229.02 and *m/z* 193.05, corresponding to losses of 44 Da (neutral loss of CO_2) and 80 Da (loss of SO_3). These findings suggested that the annotated compound was conjugated with sulfate. The other spectral family associated with the phenylpropionic scaffold also included metabolites associated with both phase I (i.e., enterolactone and hydroxyphenylpropionic acid) and phase II (i.e., dihydrocaffeic acid glucuronide) [28]. A clearer postprandial kinetics was highlighted for the molecular family of dihydroxyphenyl propanoic acid glucuronide, being addressed as phase II (r =

−0.302) metabolites [26]. Several sulfate metabolites occurred in the same molecular family, belonging to the categories of dihydrocinnamic and vanillic acids, phenolic derivatives, and indoles. Thanks to the analysis of postprandial profiles and in agreement with literature findings [29,30], the molecular scaffolds of the identified molecules are probably related to the activity of the gut microbiota. The metabolites occurring in this spectral family can be addressed to the phase I metabolism (0.526 < r < 0.700). Ultimately, it should be emphasized that some IDs belonging to the abovementioned molecular families showed structural similarities with previously annotated compounds. For example, compounds 4 and 5 of the NI dataset (**Table S2**), are characterized by retention time and MS2 fragments like those reported for the related glucuronidated conjugates found in urine by Ancillotti and co-workers [19]. In the PI dataset, twenty-four singletons were identified. In detail, several metabolites, annotated as (poly)phenolics and phenolics derivatives, were linked to phase I metabolism (e.g., dihydroxy-trimethyl-isochromenone, trihydroxybutyrophenone, and dihydroresveratrol) and with mixed contribution of phase I-II (e.g., cinnamic acid and hesperetin) by PPMC analysis. Other plant endogenous compounds, annotated with high accuracy, did not show any significant PPMC. Among them, β-glucopyranosyl-tryptophan and furaneol, as well as abscisic acid and nerol, which are well-known food-intake biomarker [31,32], and plant constituents [33,34], respectively. Some human endogenous compounds were also annotated (i.e., alpha-CEHC, ethylindole carboxylicacid, folinic acid, formylkynurenine, indole acetic acid, sebacic acid, ketodeoxycholic acid, keto-octadecadienoic acid, and hydroxy-methoxybenzophenone), exhibiting different trends against time points (−0.645 < r < 0.958), thus resulting in a complex metabolic output potentially associated with the investigated interventions, or resulting from background diet. Finally, PI mode exhibited three singletons that matched the NI annotations (i.e., azelaic acid, furoylglycine and enterolactone) and postprandial behavior interpretation based on PPMC analysis, being their longitudinal trend characterized by high and positive r-values (0.640 < r < 0.967). The other

annotated compounds occurred inside molecular families (**Table S2**), allowing for identifying interesting metabolites. **Figure 4** displays the molecular families occurring in the PI dataset with the unknown nodes labelled by "gear" symbols for which were provided hypothesized structures (**Figure S7** of the *Supplementary Materials*). The match with the PI "Nutri-Metabolomics" library identified two nodes as isomers of vanillic acid at different retention times, whereas the remaining nodes were putatively addressed as protocatechuic acid derivatives with high mass accuracy (from −3.26 to −0.06 ppm), by structural elucidation (**Figure S7**). This molecular family resulted the only one with mixed phase II-phase I contributions. In fact, vanillic acid was characterized by a statistically significant negative r-value, suggesting its direct origin from the supplements intake [35], whereas the two hypothesized protocatechuic acid derivatives exhibited an increasing signal around 6–24 h when their signals were plotted manually, probably originating from microbiota activity [30]. Most identified molecular families were related to phase I metabolism (0.441 < r < 0.930) and, interestingly, several identified and hypothesized node structures can be addressed as metabolite of the native polyphenols occurring in the bilberry and blueberry supplements [30]. Furthermore, derivatives of phloroglucinol carboxylic acid (i.e., hydroxy-dimethoxyphenyl-ethanone), cinnamic acid (i.e., coumaric acid, methylcinnamate, ferulic and isoferulic acid), and mandelic acid (i.e., methoxy-hydroxymandelate) were recognized. A deeper network inspection revealed the occurrence of in-source fragmentations of the conjugation of cinnamic acid derivatives. In detail, at the same t_R value of the compound annotated as ferulic acid (t_R = 5.14, m/z = 177.05), the feature at m/z 252.09 fragmented originating ions at m/z 177 (methoxycinnamic moiety) and at m/z 85 ($H_4SO_3+H^+$ sulfate moiety), suggesting that the annotated compound is a sulfate conjugate. Similarly, vanillic acid (t_R = 3.65, m/z = 169.05) could be addressed as sulfate conjugated, since a feature at t_R = 3.7 and m/z = 261 was characterized by fragments at m/z = 99 (H_3SO_4+) and at m/z=122 (probably benzoic acid). Finally, the compound annotated as isoferulic acid (t_R = 4.91, m/z = 177.05), coeluted with a feature at

$m/z = 263$, which is probably a derivative of dihydrocaffeic acid sulfate (annotated in NI dataset), thus supporting the sulfated conjugation of isoferulic acid.
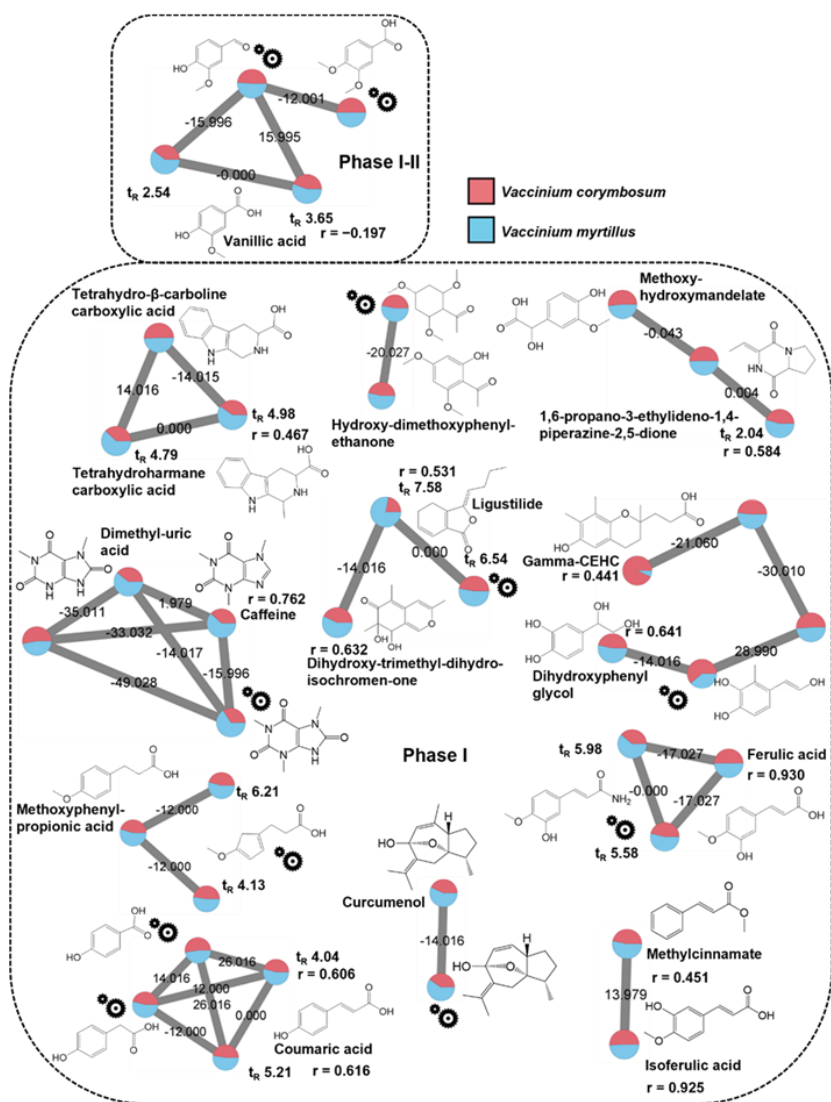
Figure 4 – Extracted networks of identified metabolites in positive ionisation mode listed in Table S2, belonging to the category of (poly)phenolic derivatives and plant endogenous constituents. Dashed boxes group the identified metabolites according to the phase I and II metabolism fol-lowing PPMC analysis. The "gear" symbols refer to the putative structure identified by manual investigation as reported in paragraph2.6 of the main text. Statistically significant Pearson cor-relation coefficients (r) are reported. Edge labels refer to the mass difference between two nodes.

Three additional interesting spectral families were identified as β-carboline derivatives (i.e., tetrahydroharmane carboxylic acid and tetrahydro-β-carboline carboxylic acid), previously identified in serum samples from this study [18], xanthine pathway metabolites (i.e., dimethyl-uric acid, caffeine), and terpene derivatives (i.e., curcumenol). Regarding xanthine derivatives, even though the identification of uric acid derivatives is in accordance with literature [36], the occurrence of caffeine has never been reported in association with berries consumption and could be attributed to the consumption of caffeine-rich foods before the fasting period foreseen in the study design and/or within the period of pool samples collection [37]. Additionally, caffeine was annotated with $\Delta = 6.1$ ppm by matching with the Massbank mass spectral library, which includes 64 spectra for caffeine acquired in heterogenous instrumental conditions. Thus, caution should be paid on this annotation. Curcumenol and its hypothesized sesquiterpene derivative were reported in this study as well as dihydroxy-trimethyl-isochromenone and ligustilide isomers (however; they were hardly related to the intake of bilberry and blueberry), as well as dihydroxy-trimethyl-isochromenone and ligustilide isomers. Finally, Gamma-CEHC, an endogenous metabolite of vitamin E [38], occurred inside molecular families, exhibiting a significant and positive r-value (0.441), representing a first report in relation to berry consumption.

6.4   Conclusions

This research investigated for the first time the applicability of the FBMN approach in combination with mass spectral libraries relevant to nutrikinetic studies as well as PPMC analysis to boost the structural annotation of postprandial urinary metabolites and to explore their nutrikinetic behavior within a two-arm intervention study on the intake of VM and VC supplements, as a relevant nutrimetabolomics application. By using the FBMN approach, 24 and 43 metabolites were annotated with high mass accuracy in NI and PI mode, respectively. The comparison with widely used annotation protocols underlined the great potential of the FBMN workflow in providing the basis for an automated

exploratory data analysis workflow resulting in a comprehensive and accurate annotation coverage. The proposed workflow offers a wider exploration of the urinary metabolome and allows for a prioritization strategy based on qualitative information. Additionally, the reliability of the presented approach was confirmed by the annotation of biochemically relevant metabolite categories across the three different annotation protocols followed. The quantitative information introduced by FBMN approach provided an estimation of the impact of the two bilberry intakes on NI and PI datasets. Furthermore, the PPMC analysis of the chromatographic areas of each identified mass feature in relation to the postprandial timepoint proved to be a successful strategy to assess the kinetic shape recognition related to phase I/phase II metabolism of IDs. It can therefore be concluded that future integration of contextual mass spectral libraries and PPMC analysis within the FBMN environment would be useful for nutrimetabolomics studies, as well as for other omics applications, where boosting annotation rates and streamlining the metabolite selection procedure are key for the data interpretation. Furthermore, it was demonstrated that the automated FBMN approach offers a versatile and scalable alternative to existing approaches that handle untargeted metabolomics profiles of biofluids for biomarker discovery. Finally, our work clearly evidenced the need for curated and contextualized mass spectral libraries that are fundamental for successful metabolite identification and thus biochemical interpretation of metabolomics profiles.

References

[1] Schrimpe-Rutledge, A.C.; Codreanu, S.G.; Sherrod, S.D.; McLean, J.A. Untargeted metabolomics strategies—Challenges and emerging directions. J. Am. Soc. Mass Spectrom. 27 (2016) 1897-1905 https://doi.org/10.1007/s13361-016-1469-y.
[2] Pomastowski, P.; Buszewski, B. Complementarity of matrix-and nanostructure-assisted laser desorption/ionization approaches. Nanomaterials, 9 (2019) 260 https://doi.org/10.3390/nano9020260.
[3] Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C.E. Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. Curr. Opin. Biotechnol. 55 (2019) 44-50 https://doi.org/10.1016/j.copbio.2018.07.010.

[4] Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinform., 8 (2007) 105 https://doi.org/10.1186/1471-2105-8-105.

[5] Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J.L. Proposed minimum reporting standards for chemical analysis. Metabolomics 3 (2007) 211-221 https://doi.org/10.1007/s11306-007-0082-2.

[6] Beniddir, M.A.; Kang, K.B.; Genta-Jouve, G.; Huber, F.; Rogers, S.; van der Hooft, J.J. Advances in decomposing complex metabolite mixtures using substructure-and network-based computational metabolomics approaches. Nat. Prod. Rep., 38 (2021) 1967-1993 https://doi.org/10.1039/D1NP00023C.

[7] van der Hooft, J.J.; de Vos, R.C.; Mihaleva, V.; Bino, R.J.; Ridder, L.; de Roo, N.; Jacobs, D.M.; van Duynhoven, J.P.; Vervoort, J. Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake. Anal. Chem., 84 (2012) 7263–7271 https://doi.org/10.1021/ac3017339.

[8] Ramos, A.E.F.; Evanno, L.; Poupon, E.; Champy, P.; Beniddir, M.A. Natural products targeting strategies involving molecular networking: Different manners, one goal. Nat. Prod. Rep., 36 (2019) 960-980 https://doi.org/10.1039/C9NP00006B.

[9] Aron, A.T.; Gentry, E.C.; McPhail, K.L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J.M.; Sikora, N.; Vargas, F. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. Nat. Protoc., 15 (2020) 1954-1991 https://doi.org/10.1038/s41596-020-0317-5.

[10] van der Hooft, J.J.; Padmanabhan, S.; Burgess, K.E.; Barrett, M.P. Urinary antihypertensive drug metabolite screening using molecular networking coupled to high-resolution mass spectrometry fragmentation. Metabolomics, 12 (2016) 125 https://doi.org/10.1007/s11306-016-1064-z.

[11] Said, I.H.; Truex, J.D.; Heidorn, C.; Retta, M.B.; Petrov, D.D.; Haka, S.; Kuhnert, N. LC-MS/MS based molecular networking approach for the identification of cocoa phenolic metabolites in human urine. Food Res. Int., 132 (2020) 109119 https://doi.org/10.1016/j.foodres.2020.109119.

[12] Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M. Feature-based molecular networking in the GNPS analysis environment. Nat. Methods, 17 (2020) 905-908 https://doi.org/10.1038/s41592-020-0933-6.

[13] Oberleitner, D.; Schmid, R.; Schulz, W.; Bergmann, A.; Achten, C. Feature-based molecular networking for identification of organic micropollutants including metabolites by non-target analysis applied to riverbank filtration. Anal. Bioanal. Chem., 413 (2021) 5291–5300 https://doi.org/10.1007/s00216-021-03500-7.

[14] Padilla-González, G.F.; Sadgrove, N.J.; Ccana-Ccapatinta, G.V.; Leuner, O.; Fernandez-Cusimamani, E. Feature-based molecular networking to target the isolation of new caffeic acid esters from yacon (Smallanthus sonchifolius,

Asteraceae). Metabolites, 10 (2020) 407 https://doi.org/10.3390/metabo10100407.

[15] Rivera-Mondragón, A.; Tuenter, E.; Ortiz, O.; Sakavitsi, M.E.; Nikou, T.; Halabalaki, M.; Caballero-George, C.; Apers, S.; Pieters, L.; Foubert, K. UPLC-MS/MS-based molecular networking and NMR structural determination for the untargeted phytochemical characterization of the fruit of Crescentia cujete (Bignoniaceae). Phytochemistry, 177 (2020) 112438 https://doi.org/10.1016/j.phytochem.2020.112438.

[16] Xie, H.-F.; Kong, Y.-S.; Li, R.-Z.; Nothias, L.-F.; Melnik, A.V.; Zhang, H.; Liu, L.-L.; An, T.-T.; Liu, R.; Yang, Z. Feature-Based Molecular Networking Analysis of the Metabolites Produced by in vitro Solid-State Fermentation Reveals Pathways for the Bioconversion of Epigallocatechin Gallate. J. Agric. Food Chem., 68 (2020) 7995-8007 https://doi.org/10.1021/acs.jafc.0c02983.

[17] Neto, F.C.; Raftery, D. Expanding Urinary Metabolite Annotation through Integrated Mass Spectral Similarity Networking. Anal. Chem., 93 (2021), 12001-12010 https://doi.org/10.1021/acs.analchem.1c02041.

[18] Renai, L.; Ancillotti, C.; Ulaszewska, M.; Garcia-Aloy, M.; Mattivi, F.; Bartoletti, R.; Del Bubba, M. Comparison of chemometrics strategies for potential exposure markers discovery and false positive reduction in untargeted metabolomics: Application to the serum analysis by LC-HRMS after intake of Vaccinium fruits supplements. Anal. Bioanal. Chem., 414 (2022), 1841-1855 https://doi.org/10.1007/s00216-021-03815-5.

[19] Ancillotti, C.; Ulaszewska, M.; Mattivi, F.; Del Bubba, M. Untargeted metabolomics analytical strategy based on liquid chromatography/electrospray ionization linear ion trap quadrupole/orbitrap mass spectrometry for discovering new polyphenol metabolites in human biofluids after acute ingestion of vaccinium myrtillus berry supplement. J. Am. Soc. Mass Spectrom., 30 (2019) 381-402 https://doi.org/10.1007/s13361-018-2111-y.

[20] Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinform., 11 (2010) 395 https://doi.org/10.1186/1471-2105-11-395.

[21] Vargas, F.; Weldon, K.C.; Sikora, N.; Wang, M.; Zhang, Z.; Gentry, E.C.; Panitchpakdi, M.W.; Caraballo-Rodríguez, A.M.; Dorrestein, P.C.; Jarmusch, A.K. Protocol for community-created public MS/MS reference spectra within the Global Natural Products Social Molecular Networking infrastructure. Rapid Commun. Mass Spectrom., 34 (2020) e8725 https://doi.org/10.1002/rcm.8725.

[22] Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res., 13 (2003) 2498-2504 https://doi.org/10.1101/gr.1239303.

[23] Pellis, L.; van Erk, M.J.; van Ommen, B.; Bakker, G.C.; Hendriks, H.F.; Cnubben, N.H.; Kleemann, R.; van Someren, E.P.; Bobeldijk, I.; Rubingh, C.M. Plasma metabolomics and proteomics profiling after a postprandial challenge

reveal subtle diet effects on human metabolic status. Metabolomics, 8 (2012) 347–359 https://doi.org/10.1007/s11306-011-0320-5.

[24] Garcia-Aloy, M.; Ulaszewska, M.; Franceschi, P.; Estruel-Amades, S.; Weinert, C.H.; Tor-Roca, A.; Urpi-Sarda, M.; Mattivi, F.; Andres-Lacueva, C. Discovery of intake biomarkers of lentils, chickpeas, and white beans by untargeted LC–MS metabolomics in serum and urine. Mol. Nutr. Food Res., (2020) 64 1901137 https://doi.org/10.1002/mnfr.201901137.

[25] Iglesias-Carres, L.; Mas-Capdevila, A.; Bravo, F.I.; Aragonès, G.; Arola-Arnal, A.; Muguerza, B. A comparative study on the bioavailability of phenolic compounds from organic and nonorganic red grapes. Food Chem., 299 (2019) 125092 https://doi.org/10.1016/j.foodchem.2019.125092.

[26] Ancillotti, C.; Ciofi, L.; Rossini, D.; Chiuminatto, U.; Stahl-Zeng, J.; Orlandini, S.; Furlanetto, S.; Del Bubba, M. Liquid chromatographic/electrospray ionization quadrupole/time of flight tandem mass spectrometric study of polyphenolic composition of different Vaccinium berry species and their comparative evaluation. Anal. Bioanal. Chem., 409 (2017) 1347-1368 https://doi.org/10.1007/s00216-016-0067-y.

[27] de Mello, V.D.; Lankinen, M.A.; Lindström, J.; Puupponen-Pimiä, R.; Laaksonen, D.E.; Pihlajamäki, J.; Lehtonen, M.; Uusitupa, M.; Tuomilehto, J.; Kolehmainen, M. Fasting serum hippuric acid is elevated after bilberry (Vaccinium myrtillus) consumption and associates with improvement of fasting glucose levels and insulin secretion in persons at high risk of developing type 2 diabetes. Mol. Nutr. Food Res., 61 (2017) 1700019 https://doi.org/10.1002/mnfr.201700019.

[28] Vetrani, C.; Rivellese, A.A.; Annuzzi, G.; Adiels, M.; Borén, J.; Mattila, I.; Orešič, M.; Aura, A.-M. Metabolic transformations of dietary polyphenols: Comparison between in vitro colonic and hepatic models and in vivo urinary metabolites. J. Nutr. Biochem., 33 (2016) 111-118 https://doi.org/10.1016/j.jnutbio.2016.03.007.

[29] Aravind, S.M.; Wichienchot, S.; Tsao, R.; Ramakrishnan, S.; Chakkaravarthi, S. Role of dietary polyphenols on gut microbiota, their metabolites and health benefits. Food Res. Int., 142 (2021) 110189 https://doi.org/10.1016/j.foodres.2021.110189.

[30] Stevens, J.F.; Maier, C.S. The chemistry of gut microbial metabolism of polyphenols. Phytochem. Rev., 15 (2016) 425-444 https://doi.org/10.1007/s11101-016-9459-z.

[31] Du, X.; Finn, C.E.; Qian, M.C. Volatile composition and odour-activity value of thornless 'Black Diamond' and 'Marion' blackberries. Food Chem., 119 (2010) 1127-1134 https://doi.org/10.1016/j.foodchem.2009.08.024.

[32] Gutsche, B.; Grun, C.; Scheutzow, D.; Herderich, M. Tryptophan glycoconjugates in food and human urine. Biochem. J., 343 (1999) 11-19 https://doi.org/10.1042/bj3430011.

[33] Degu, A.; Ayenew, B.; Cramer, G.R.; Fait, A. Polyphenolic responses of grapevine berries to light, temperature, oxidative stress, abscisic acid and jasmonic acid show specific developmental-dependent degrees of metabolic

resilience to perturbation. Food Chem., 212 (2016) 828-836 https://doi.org/10.1042/bj3430011.

[34] Elsharif, S.A.; Buettner, A. Structure–odor relationship study on geraniol, nerol, and their synthesized oxygenated derivatives. J. Agric. Food Chem., 66 (2016) 2324-2333 https://doi.org/10.1021/acs.jafc.6b04534.

[35] Colak, N.; Primetta, A.K.; Riihinen, K.R.; Jaakola, L.; Grúz, J.; Strnad, M.; Torun, H.; Ayaz, F.A. Phenolic compounds and antioxidant capacity in different-colored and non-pigmented berries of bilberry (Vaccinium myrtillus L.). Food Biosci., 20 (2017) 67-78 https://doi.org/10.1016/j.fbio.2017.06.004.

[36] Lotito, S.B.; Frei, B. Consumption of flavonoid-rich foods and increased plasma antioxidant capacity in humans: Cause, consequence, or epiphenomenon? Free Radic. Biol. Med., 41 (2006) 1727-1746 https://doi.org/10.1016/j.freeradbiomed.2006.04.033.

[37] Martínez-López, S.; Sarriá, B.; Gómez-Juaristi, M.; Goya, L.; Mateos, R.; Bravo-Clemente, L. Theobromine, caffeine, and theophylline metabolites in human plasma and urine after consumption of soluble cocoa products with different methylxanthine contents. Food Res. Int., 63 (2014) 446-455 https://doi.org/10.1016/j.foodres.2014.03.009.

[38] Wu, J.; Croft, K. Vitamin E metabolism. Mol. Asp. Med., 28 (2007) 437-452 https://doi.org/10.1016/B978-0-12-811907-5.00020-8.

7 Development of a comprehensive two-dimensional liquid chromatographic mass spectrometric method for the non-targeted identification of poly- and perfluoroalkyl substances in aqueous film-forming foams.

Supplementary Materials:

https://ars.els-cdn.com/content/image/1-s2.0-S000326702201056X-mmc1.docx

Abstract

In this research, we developed an online comprehensive two-dimensional liquid chromatographic (LC×LC) method hyphenated with high-resolution mass spectrometry (HRMS) for the non-targeted identification of poly- and perfluorinated compounds (PFASs) in fire-fighting aqueous-film forming foams (AFFFs). The method exploited the combination of mixed-mode weak anion exchange-reversed phase with an octadecyl stationary phase, separating PFASs according to ionic classes and chain length. To develop and optimize the LC×LC method we used a reference training set of twenty-four anionic PFASs, representing the main classes of compounds occurring in AFFFs and covering a wide range of physicochemical properties. In particular, we investigated different modulation approaches to reduce injection band broadening and breakthrough in the second dimension separation. Active solvent and stationary phase assisted modulations were compared, with the best results obtained with the last approach. In the optimal conditions, the predicted peak capacity corrected for undersampling was higher than three-hundred in a separation space of about 60 min. Subsequently, the developed method was applied to the non-targeted analysis of two AFFF samples for the identification of homologous series of PFASs, in which it was possible to identify up to thirty-nine potential compounds of interest utilizing Kendrick mass defect analysis. Even within the samples, the features considered potential PFAS by mass defect analysis elute in the chromatographic regions discriminating for the ionic group and/or the chain length, thus confirming the applicability of the method presented for the analysis of AFFF mixtures and, to a further extent, of environmental matrices affected by the AFFF.

Keywords: orthogonality, peak capacity, modulation, data processing, feature detection, mass defect.

7.1   Introduction

The application of online comprehensive two-dimensional liquid chromatography (LC×LC) to the analysis of complex samples represents an attractive analytical

solution due to the increase of peak capacity (i.e., number of peaks in the separation space) reached by the combination of two different separation mechanisms (i.e., differences in selectivity) [1,2] and thus capable to extend the knowledge on sample composition. Nowadays, the most common hardware configuration adopted to transfer fractions automatically from the first to the second chromatographic dimension is a 4 ports duo valve (also named as modulator) which interfaces the two columns. Starting from this, other valve-based modulation systems have been proposed, such as multiple heart-cutting valve and 2-position 10-port valves, additionally combined with modulation-assistant technology to increase the compatibility between the two dimensions [3]. When the fractions are collected during the entire first dimension chromatographic run, it is referred as "comprehensive" LC×LC [2]. This technique has already found several applications in the analysis of synthetic and natural polymeric mixtures, natural compounds, metabolomics, food analysis, and biomolecules (e.g., antibodies, intact and digested proteins) [4-8], proving its effectiveness in increasing the chromatographic resolution compared to mono-dimensional (1D) LC by pushing system peak capacity of the chromatographic system up to 1000. Besides these successful applications, LC×LC found limited interest in the development of methods in environmental analytical chemistry, especially in non-targeted mode [9], due to its complex management and the hampered sensitivity during each modulation, making the method not applicable to the detection of contaminants at trace levels (ng L^{-1} or sub ng L^{-1}) [10,11]. Per- and polyfluoroalkyl substances (PFASs) are anthropogenic chemicals with unique physicochemical properties, which have been widely used since the1950s for many industrial and commercial applications, as well as in the productions of fluoropolymers [12]. Among their various applications, PFASs are relevant components of aqueous film-forming foams (AFFFs), i.e. fire-extinguishing reagents for hydrocarbon fuel-based fires that prevent from reignition by reducing surface tension and oxygen transfer [13,14]. PFAS structural uniqueness makes them hard to replace in these applications, and at the same time generates a

growing environmental concern, since their thermal and chemical stability, low vapor pressure and high solubility, make PFASs persistent and mobile chemicals [15]. Throughout the years, a great focus on the environmental impact on anionic PFASs has been made, promoting their inclusion in several regulations, such as the Stockholm Convention on Persistent Organic Pollutants (2009) [16], the 1907/2006/EU (i.e., list of candidate substances to be regulated) [17], the Directive 2013/39/EU (i.e., priority substances in the field of water policy) [18], and most recently in the Directive 2020/2184/EU on water intended for human consumption [19]. Thus, the characterization of AFFF blends represents a key point for the assessment of the environmental impact of PFASs. Since AFFF are complex mixtures of fluorinated and non-fluorinated surfactants, solvents, and additives [20], their analysis has been performed using non-targeted platforms based on 1D-LC hyphenated with high resolution mass spectrometry (HRMS). In brief, several homologous series of anionic, cationic and zwitterionic PFASs have been identified inside new-generation and legacy AFFF mixtures (e.g., 3M Light water), such as perfluoroalkyl sulfonic acids, perfluoroalkylsulfonamides, fluorotelomer betaines, fluorotelomer thioethers, fluorotelomer sulfonamides and fluorotelomer amino carboxylic acids [14,21,22]. However, these studies have been carried out on octadecyl (C18) stationary phases, thus limiting the characterization of the analyzed mixtures to PFASs with $logD_{pH=7}>0$ (i.e., containing a minimum number of carbon atoms equal to six), while those with $logD_{pH=7}<0$, known to be present in the investigated AFFF mixtures, were not detected probably due to column breakthrough. Moreover, PFAS detection in the investigated samples was performed by mass domain deconvolution, due to the insufficient chromatographic resolution provided by C18 columns, which gave rise to the co-elution of PFASs characterized by different polar functional groups and the same carbon chain length. A first attempt to obtain a "class-dependent" elution order of PFASs to improve their identification rate in AFFF mixtures, was carried out by implementing an orthogonal LC separation based on weak cation and weak anion exchange guard columns in series with a C18 column [23]. Under

these conditions, the retention of zwitterionic and cationic PFASs was increased in comparison with the group of anionic homologues, although the chromatographic overlap was not completely resolved. The need to obtain a suitable chromatographic separation for the identification of PFAS classes in complex samples, combined with the trend dictated by the regulation to use smaller (<C8) and more polar PFASs (i.e., less bioaccumulative) in modern products, led to the use of alternative stationary phases to C18. In this regard, targeted 1D-LC-MS platforms using mixed-mode reversed phase weak anion exchange (RPWAX) columns provided promising results in detection and quantitation of short-chain PFASs (e.g., perfluoropropanoic acid, perfluoromethan sulfonic acid, and trifluoroacetic acid), while being capable of retaining less polar PFASs [24-26]. Moreover, the mixed-mode chemistry allows to exploit pH and ionic strength of the mobile phases to increase retention (by the ion exchange mechanisms) as a function of the charged polar group (e.g., carboxylate vs sulfonate) occurring in the investigated analytes, hindering part of the hydrophobic partitioning typical of reversed phase chemistry [27].

Based on these considerations, the aim of this research was to develop a novel non-targeted and comprehensive LC×LC -HRMS method for the analysis of PFASs in AFFF mixtures, exploiting the selectivity of the combination of RPWAX×C18 stationary phases, which until now has been used only for the two-dimensional analysis of small biomolecules [28,29]. In particular, the aim is to obtain a separation in the two-dimensional space according to (i) the classes of PFASs (i.e., ionic groups inside the molecules) and (ii) their chain length, allowing for increasing the detection rate by exploiting the two-dimensional separation space and reducing the complexity in the mass domain. The developed method is then applied to the analysis of AFFF samples, thus laying the groundwork for its future application on AFFF-impacted environmental matrices.

7.2 Materials and methods

7.2.1 Chemicals and reagents

Acetonitrile (ACN, MS grade), methanol (HPLC grade) and water (MS grade) were purchased from Biosolve Chimie (Dieuze, France). Milli-Q water (18.2 MΩ·cm) was obtained from a purification system (Millipore, Bedford, MA, USA). Ammonium acetate (AcNH4, analytical grade; 98%) and acetic acid (analytical grade; 99%) were purchased from Sigma-Aldrich (Darmstadt, Germany). The "Perfluorinated Native Compound Standard" mixture (PFC-24-10X) adopted as the analyte training set for method development, was purchased by AccuStandard (New Haven, CT, USA), and its details and composition are reported in **Table S1** of the *Supplementary Materials*. Perfluoro-n-[1,2-^{13}C2]octanoic acid (M2PFOA, 2 mg L^{-1}) and perfluoro-n-[1,2,3,4-^{13}C4]octansulfonate (MPFOS, 2 mg L^{-1}) used as internal standards were purchased from Wellington Laboratories (Ontario, Canada). The analytical columns tested for method development were purchased as follow: Zorbax Eclipse Plus C18 (2.1 x 50 mm, 1.8 µm) from Agilent (Santa Clara, CA, USA), Atlantis Premier BEH C18 AX (2.1 x 30 mm, 1.7 µm) from Waters (Millford, MA, USA), and Acclaim WAX-1 (3 x 50 mm, 3 µm) from ThermoFisher Scientific (Waltham, MA, USA). The delay column (Zorbax Eclipse Plus C18, 2.1 x 5 mm, 1.8 µm) was obtained from Agilent. The equipment used for the modulation valve of the LC×LC system consisted of a capillary (0.12 x 120 mm, Agilent Technologies, Waldbronn, Germany) and two C18 guard columns (SecurityGuard Ultra AJ0–87822.1×2 mm and sub-2 µm particle, Phenomenex, Utrecht, NL) used for analyte trapping.

7.2.2 Samples

Aliquots of 1.2 mL of the two native AFFF mixtures investigated in this study were collected from bulk products purchased as follows. 3M Light Water 3% (3M) fire-fighting foam was purchased from 3M (Minnesota, USA) and Orchidex 3% Eco (ORC) was obtained from Orchidee (Tisselt, Belgium). For

samples analysis, the investigated AFFF mixture were diluted 1:100 with a methanolic solution (water/methanol 80/20, v/v%). Moreover, pooled samples were prepared by mixing equal aliquots of 3M and ORC, maintaining the same dilution factor. Before analysis, individual and pooled samples (1 mL aliquots each) were spiked with 50 µL of the 1 mg L^{-1} internal standard mixture (i.e., M2PFOA and MPFOS) to the final concentration of 10 µg L^{-1}.

7.2.3   Instrumental conditions

The potential PFAS contamination by the system was prevented by using a C18 delay column downstream of the mixers in the respective 1D-LC and LC×LC configurations. **Figure S3** shows the acquisition of a blank in the operating conditions of the first dimension. LC-HRMS experiments for the evaluation of orthogonality for among the tested stationary phases were performed on ACQUITY UPLC Quaternary Solvent Manager and ACQUITY Sample Manager (Waters), coupled with a Bruker microTOF-Q mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization (ESI) source. **Table S2** in section S2 of the *Supplementary Materials* reports the chromatographic setups adopted. The MS acquisitions were performed using a mass range of 100-1000 *m/z* and a scan rate of 1 Hz (focus mode active ad an estimated resolution of 35k). Data were acquired in negative ionization mode using a capillary voltage of 3.5 kV, a coaxial nebulizer gas flow of 9.0 L min^{-1} at 200° C and 3.0 bar of pressure. The PFASs standard mixture was diluted to reach the concentration value of 0.5 mg L^{-1} with a water/methanol solution (80/20, v/v%) and the injection volume was 10 µL. LC×LC experiments were carried out on an Agilent Infinity LC×LC system (Agilent, Waldbronn, Germany), which included an autosampler (G4226A) equipped with a 50 µL loop, two binary pumps (G7120A), one isocratic pump (G1310A), a thermostatted column oven (G1316A), and a modulation valve system (G1170A) used for active solvent modulation (ASM) and stationary phase assisted modulation (SPAM), which is graphically shown in **Figure S4** of the *Supplementary Materials*. **Table 1** reports the setups that were investigated to develop the LC×LC method, being the setup

6 the one adopted for sample analyses. Briefly, 1D chromatography was performed on the Atlantis C18 AX analytical column, adopting a flow rate of 0.1 mL min^{-1} and AcNH4 20mM Milli-Q water (A1) and 90% of ACN (B1) as mobile phases, both adjusted at pH 4 with acetic acid. A gradient elution was adopted, running for 50 minutes starting from 25% of B1, before re-equilibrating the system for the subsequent injection (total analysis time equal to 60 min). LC×LC chromatography was carried out using the Zorbax Eclipse Plus C18 analytical column with a flow rate of 0.7 mL min^{-1} using 2mM AcNH4 MS grade water (A2) and ACN (B2) as eluents. The gradient program was set at 0.46 min starting from the 10% of B2, on an overall time for LC×LC analysis of 0.55 minutes, i.e. equal to the adopted modulation time. The make-up solvent for SPAM configuration was a 2mM AcNH4 solution in Milli-Q water, which was dispensed at a flow rate of 1 mL min^{-1}. Injection volume was set at 20 µL for the injection of PFASs standard mixture (1 mg L^{-1} in water/methanol solution 80/20, v/v%) and AFFF diluted samples. The LC×LC system was coupled with an Orbitrap Q Exactive Plus from ThermoFisher Scientific by an Heated Electrospray Probe (HESI-II). Full scan analyses in negative ionization mode were performed in the 80-1000 *m/z* range using a resolution of 35000 at 200 *m/z*. The ESI source parameters were set as follow: sheat gas=60, auxiliary gas=20, spray voltage=2.5 kV, capillary temperature=380° C, S-lens RF level=50, and auxiliary gas heater temperature=350° C.

7.2.4   *Orthogonality evaluation of the RPWAX-C18 stationary phases*

Scatter plots between the RPWAX and C18 data were obtained using the retention times (t_R) of the compounds inside PFAS training set, which were extracted using the [M-H]- pseudo-molecular ion masses (i.e., XICs) reported in **Table S1**. The t_Rs of the occurring PFASs were normalized (normalized retention time, NRT) using the following equation.

$$NRT = \frac{t_{Ri} - t_{Rmin}}{t_{Rmax} - t_{Rmin}} \qquad (1)$$

Where t_{Rmin} and t_{Rmax} represent the first and the latest eluting peak in PFAS training set, respectively, and t_{Ri} is the retention time of the chosen peak. Orthogonality was estimated by linear regression fitting of NRTs from each investigate setup, carrying out on Excel the "Regression" macro from "Data analysis" component, using the obtained correlation coefficient (R^2) as an orthogonality degree estimator (i.e., the closer to 1, the lower the degree of orthogonality).

7.2.5   *Analysis of LC×LC data for method development and sample analysis*

The investigation on the development of the LC×LC-HRMS method was carried out using the training set of PFASs (**Table S1**) and by processing the acquired data under the investigated LC×LC setups (see **Table 1**) with a two-dimensional (2D) feature detection algorithm, described in the following lines, to obtain the scatter plots of the two-dimensional chromatographic space. Briefly, the 2D feature detection algorithm was developed using the julia (1.7.1) programming language and processing the acquired data with the Self-Adjusting Feature Detection (SAFD) algorithm [30], after the conversion into .mzXML open format via MSConvert provided by the ProteoWizard package (section S3 of the *Supplementary Materials*) [31]. Subsequently, the feature list obtained from SAFD was processed with the implemented "grouping" script to obtain a new list containing the identified 2D features, which are described by their exact mass, retention time in first dimension (t_{R1D}), retention time in second dimension (t_{R2D}), maximum intensity, and the number of modulations in which they occur. The obtained 2D feature list was used to build the scatter plots (t_{R1D} vs t_{R2D}) representing the two-dimensional separation space. The full details on data pre-treatment and 2D feature detection algorithm are reported in section S3 of the *Supplementary Materials*. The separation obtained within each setup for method development was evaluated by calculating the predicted peak capacity ($n'_{c,2D}$), using the equation proposed by Li et al. [32], as described in section S4 of the *Supplementary Materials*.

The 2D feature detection algorithm described above was used also to analyze the non-targeted LC×LC-HRMS data of the investigated samples. To identify the potential PFAS homologous series occurring in 3M and ORC samples, the Kendrick mass defect (KMD) was applied as elsewhere reported [33,34]. Briefly, the exact mass of the detected 2D features were converted to the mass scale related to the mass difference of 50 Da (i.e., -CF2-). Then, the mass defects were calculated by subtracting the converted nominal mass (rounded down) from the converted exact masses. To obtain mass defect plots, the *m/z* values (IUPAC scale) were plotted against the mass defect values. Inside each plot, two distinct mass defect regions were selected for the evaluation of potential PFASs occurrence, i.e. 0.0-0.1 and 0.9-1.0. According to non-targeted protocols [35], the limits of identification related to the trining set of PFASs was evaluated as the concentration corresponding to the signal threshold used for the feature detection (i.e., minimum of detectable ion intensity). Principal component analysis (PCA) was used to evaluate data quality of the investigated real samples by using the intensities of the detected features after raw data alignment by MZmine 2 (http://mzmine.github.io/). PCA was performed using the function prcomp() of the package *FactoMineR* in R environment (https://www.r-project.org/).

7.3 Results and discussion

The development of the non-targeted LC×LC-HRMS method was carried out in two main steps, the first focused on the study of the orthogonality of different stationary phases, including RPWAX materials, and the second focused on the two-dimensional RPWAX×RPLC method optimization. A training set comprising twenty-four AFFF relevant PFAS (see **Table S1** and **Figure S1**) was used for method development. In detail, the classes of PFASs present in the training set were: perfluoroalkyl carboxylic acids (PFACAs), perfluoroalkyl sulfonic acids (PFASAs), N-methylperfluorosulfonamidoacetic acids (Me-FSAAs), N-ethylperfluorosulfonamidoacetic acids (Et-FSAAs), fluorotelomer sulfonic acids (FTSs), and perfluoroalkyl sulfonamides (PFSAs), thus covering a wide range of physicochemical properties. The proposed RPWAX×RPLC method separates the

analytes in the training set in the two-dimensional space according to analyte ionic groups and the length of the alkyl chain. Finally, the developed method is applied to the non-targeted analysis of the two AFFF samples (see paragraph 2.2), in which the detected features are analyzed by means of KMD for PFASs identification.

7.3.1 RPWAX-C18 orthogonality tests

Aiming to develop a separation combining orthogonal selectivities, we compared the orthogonality of different 1D RPWAX and C18 based methods, different RPLC×RPLC setups were preliminarily tested, the plots of which are shown in **Figure S1** of the *Supplementary Materials*. Linear regression approach was used as metric to compare the correlation between the two separations. **Figs. S2A** and **S2B** show the combination between the C18 column used in this study (**Table S2**) and a column characterized by the same particle size but a different bonding chemistry of the octadecyl functions (i.e., ethylene bridged bonds), which increases the chromatographic efficiency and stability with respect to pH. These setups were tested in two different ion-paring conditions to reduce the influence of the ionic group in the reversed phase separation, in the first case by adding formic acid (0.1%) and in the second with diethylamine (10 mM, adjusting the pH at 7), and compared to neutral conditions on the C18 column. In both cases a strong coelution of the homologous series was observed, making these conditions too correlated with each other (R2 of 0.7-0.9). A comparable result was observed in **Fig. S2C**, using a pentafluoro-phenyl column, which introduces hydrogen bond and dipole-dipole interactions into the separation mechanism in addition to the hydrophobic contribution. These results confirm that the RPLC×RPLC combination is not enough orthogonal and thus suitable for the separation of homologous series of PFASs as a function of the ionic groups.

To introduce a separation based on a different selectivity that however maintains good compatibility with RPLC we tested mixed mode separations. Two types of RPWAX stationary phases were tested, i.e. Acclaim WAX-1 and Atlantis C18

AX. The first one is based on ethylene-bridged hybrid particles containing both C18 groups and tertiary alkylamine moieties (pKa~8-9) bonded separately, whereas the second one is characterized by bonded groups that incorporate both hydrophobic (octadecyl) and anionic moieties (amide and tertiary amine functions) together in the same structure [36]. The methods conditions summarized in **Table S2** were investigated to evaluate the orthogonality in respect of the training set of PFASs (see **Table S1**), thus simulating their behaviour inside AFFF mixtures. **Figure 1** reports the scatter plots and the regression trend lines of the four setups, also highlighting the R^2 of each regression model. As a general overview, RPWAX stationary phases exploited the partitioning of the analytes according to the polar group occurring in each compound structure, i.e. carboxylates (mainly above the trend line, i.e. late eluting on the RPWAX dimension), sulfonates, and sulfonamides (below the trend line, i.e. early eluting on the RPWAX dimension), along with the elution order related to hydrophobic interactions in respect of PFAS chain length. This trend highlights how mixed-mode interactions help increasing the orthogonality of PFAS separation in combination with C18 (hydrophobic interactions only), which otherwise would have been too correlated (**Fig. S2**). **Figs. 1A** (setup 1 vs 2) and **1D** (setup 1 vs 5) show a similar orthogonality scenario, being indeed characterized by very close R^2 values (~0.68), suggesting that the combination of a "salt" gradient (i.e., increasing the percentage of AcNH4) in time with the organic fraction gradient did not provide a significant increase of the orthogonality. On the same RPWAX stationary phase, a slight improvement in orthogonality was obtained by increasing the initial percentage of organic up to 65% which allowed for reducing the NRT correlation (i.e., $R^2 = 0.67$), as displayed in **Fig. 1C**. Simultaneously with the damping of hydrophobic interactions due to the increase in organic strength, this outcome can be addressed to the decreased water content in the mobile phase, which reduced the degree of hydration of the investigated surfactants, i.e. allowing for increasing anionic exchange interactions by reducing the hydration of the polar group inside each analyte [37]. Even if the R^2 value increased up to 0.88 and the

mixed-mode interactions seemed to be weaker on the Atlantis C18 AX, **Fig. 1B** showed a relevant change in the retention of PFBuA and the others short-chain PFASs (i.e., PFPeA, PFBuS, PFHxA, and Hx-FTS), being more spaced from long-chain homologues in RPWAX dimension, and thus potentially increasing the peak coverage in the first dimension in the 2D-LC combination. For these reasons, conditions 1 vs 3 (**Fig. 1B**) and 1 vs 4 (**Fig. 1C**) were selected for the optimization of the 2D-LC configuration reported in the next paragraphs.
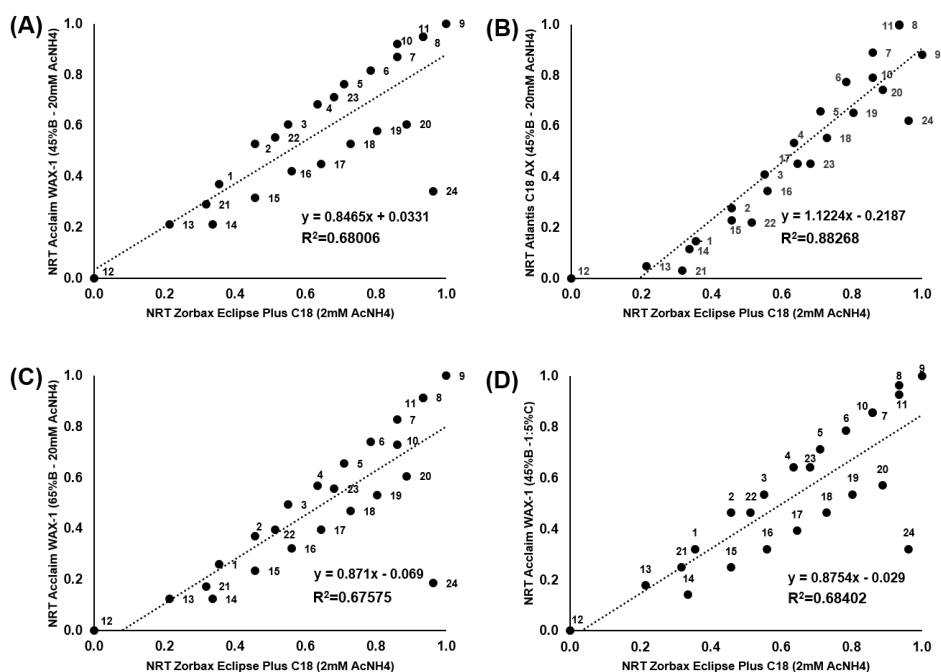
Figure 1 − Normalized retention time (NRT) scatter plots and regression trend lines of setups (A) 1 vs 2, (B) 1 vs 3, (C) 1 vs 4, and (D) 1 vs 5, investigated for the evaluation of the system orthogonality as reported in Table S2, using the PFAS training set. The equation of the regression trend line and the relative correlation coefficient (R^2) are also displayed in each plot. 1=PFHxA, 2=PFHpA, 3=PFOA, 4=PFNA, 5=PFDA, 6=PFUnDA, 7=PFDoDA, 8=PFTrDA, 9=PFTeDA, 10=Me-FSA, 11=Et-FSA, 12=PFBuA, 13=PFPeA, 14=PFBuS, 15=PFPeS, 16=PFHxS, 17=PFHpS, 18=PFOS, 19=PFNS, 20=PFDS, 21=Hx-FTS, 22=Oc-FTS, 23=D-FTS, 24=PFOSA. Abbreviation meanings are reported in Table S1 of the Supplementary Materials.

7.3.2   LC×LC method development

7.3.2.1   Comparison of RPWAX columns in LC×LC

Table 1 reports the LC×LC methods investigated for the analysis of PFASs in AFFF mixtures. Since both the selected setups from the paragraph 3.1 are characterized by high organic strength in the first dimension (e.g., 45% and 65% ACN used at the beginning of the gradient), and the injection of these eluents in the second separation dimension (30 µL from the 1D are injected, about 25% of the 2D column volume) can cause peak distortions, breakthrough and reduce chromatographic resolution [38], our study evaluated at first ASM (**Figure S4A**) approaches to reduce the contribution of these effects . **Figure 2** shows the scatter plots of the two-dimensional space for the investigated 2D-LC setups using the training set of the twenty-four PFASs. In addition, **Figures S6A-S6F** of *Supplementary Materials* show the contourplots of the data acquired in the investigated LC×LC setups. The scatter plot related to setup 1 is shown in **Fig. 2A**, in which a full gradient method is applied to the combination of Acclaim WAX-1 in the first dimension and the Zorbax Eclipse Plus C18 in the second dimension, adopting a modulation time of 1 min and an ASM×2 factor for 0.14 min. The wider internal diameter of the Acclaim WAX-1 column determines a larger column volume (i.e., 3 mm, no smaller ID column is currently available by the vendor) which cause a high gradient delay at 0.03 mL min^{-1} (flow rate used in 2DLC experiments) and requires higher gradient volumes respect to 2.1 mm ID columns. Therefore to compensate for this, a steep organic strength gradient (i.e., a short gradient time) was adopted to ensure the elution of the entire training set, over a total analysis time of 120 min. These conditions generated a strong coelution in the first dimension, which hindered the ionic group selectivity, resulting

Table 1 − Stationary and mobile phases, flow rates (mL min^{-1}), gradient program, modulation time (MT, min), and configured modulation system (CMS) used in the investigated 2D-LC setups for the method development. ^{1}D=first dimension, ^{2}D=second dimension. Active solvent modulation (ASM) conditions are reported as ratio/time, whereas for stationary phase assisted modulation (SPAM) the flow rate (mL/min) of the make-up eluent is reported. In all the setups, the ^{2}D column was thermostatted at 50° C.

Setups	Analytical Columns	Mobile Phases	Flow rate	Gradient program	MT	CMS
1	^{1}D: Acclaim WAX-1	A1: Water 20mM AcNH4 (pH 5.5) B1: ACN:H2O 9:1 20mM AcNH4	0.1 (0-2min) 0.03 (2-120min)	0-3min - 65%B 3-13min - 100%B 13-95min - 100%B 95-100min - 65%B 100-120min - 65%B		ASM 1:2/0.14 min
	^{2}D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.15min - 10%B 0.15-0.70min - 100%B 0.70-0.80min - 100%B 0.80-0.81min - 10%B 0.81-1.00min - 10%B	1	
2	^{1}D: Atlantis C18 AX	A1: Water 20mM AcNH4 (pH 5.5) B1: ACN:H2O 9:1 20mM AcNH4	0.03	0-3min - 45%B 3-40min - 100%B 40-50min - 100%B 50-50.50min - 45%B 50.50-60min - 45%B		ASM 1:2/0.14 min
	^{2}D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.15min - 10%B 0.15-0.70min - 100%B 0.70-0.80min - 100%B 0.80-0.81min - 10%B 0.81-1.00min - 10%B	1	
3	^{1}D: Atlantis C18 AX	A1: Water 20mM AcNH4 (pH 5.5) B1: ACN:H2O 9:1 20mM AcNH4	0.03	0-3min - 45%B 3-40min - 100%B 40-50min - 100%B 50-50.50min - 45%B 50.50-60min - 45%B		ASM 1:2/0.14 min
	^{2}D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.15min - 10%B 0.15-0.20min - 45%B 0.20-0.65min - 65%B 0.65-0.70min - 100%B 0.70-0.76min - 100%B 0.76-0.80min - 10%B 0.80-1.00min - 10%B	1	

4	[1]D: Atlantis C18 AX	A1: Water 20mM AcNH4 (pH 4) B1: ACN:H2O 9:1 20mM AcNH4 (pH 4)	0.03	0-3min - 35%B 3-29min - 100%B 29-50min - 100%B 50-50.50min - 35%B 50.50-60min - 35%B	1	SPAM 0.3 mL/min
	[2]D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.09min - 10%B 0.09-0.20min - 45%B 0.20-0.65min - 65%B 0.65-0.70min - 100%B 0.70-0.76min - 100%B 0.76-0.80min - 10%B 0.80-1.00min - 10%B		
5	[1]D: Atlantis C18 AX	A1: Water 20mM AcNH4 (pH 4) B1: ACN:H2O 9:1 20mM AcNH4 (pH 4)	0.1	0-1min - 25%B 1-51min - 100%B 51-55min - 100%B 55-55.50min - 35%B 55.50-60min - 35%B	0.55	SPAM 1.0 mL/min
	[2]D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.03min - 10%B 0.03-0.08min - 45%B 0.08-0.27min - 65%B 0.27-0.35min - 100%B 0.55-0.40min - 100%B 0.40-0.41min - 10%B 0.41-0.55min - 10%B		
6	[1]D: Atlantis C18 AX	A1: Water 20mM AcNH4 (pH 4) B1: ACN:H2O 9:1 20mM AcNH4 (pH 4)	0.1	0-1min - 25%B 1-51min - 100%B 51-55min - 100%B 55-55.50min - 35%B 55.50-60min - 35%B	0.55	SPAM 1.0 mL/min
	[2]D: Zorbax Eclipse Plus C18	A2: 2mM AcNH4 B2: ACN 2mM AcNH4	0.7	0.00-0.02min - 10%B 0.02-0.08min - 45%B 0.08-0.27min - 65%B 0.27-0.45min - 100%B 0.45-0.48min - 100%B 0.48-0.49min - 10%B 0.49-0.55min - 10%B		

in two major elution bands, the first one related to sulfonate compounds ($t_{R1D} \sim 45$ min) and second one to carboxylate PFASs ($t_{R1D} \sim 50$ min). On the other hand, even when ASM was adopted, the high organic strength of the mobile phase from the first dimension led to the breakthrough of the short-chain and more polar analytes (e.g., PFBuA, PFPeA, and PFBuS), as illustrated in the **Figures S5A1**, **S5B1**, and **S5C1** of the *Supplementary Materials*. **Fig. 2B** reports the scatter plot obtained from setup 2, using the Atlantis C18 AX column in the first dimension. The advantages of this setup are related to the first dimension method and are the (i) different dimensions of the column (i.e., 2.1 mm of internal diameter and 1.7 µm of particle size), the (ii) lower organic strength and the (iii) shallower gradient adopted in the first dimension method (**Table 1**). These conditions allowed to effectively use more of the 1D separation space in the analysis of the training set of PFASs, in comparison with the setup 1 (elution over 56.7 min over 60 min in setup 2 with respect to 64.6 min over 120 min in setup 1), resulting in an increased peak capacity ($n'_{c,2D}$=80) of the LC×LC system and in a reduced undersampling correction factor (1/β) (**Table S4** of the *Supplementary Materials*). Additionally, the lower percentage of organic modifier in the mobile phase made more effective the contribution of ASM, which increased the retention of the more polar analytes in the second dimension, making also less prominent the peak splitting on the second dimension (**Figures S5A2**, **S5B2**, and **S5C2**).

7.3.2.2   *LC×LC modulation and method optimization*

Although the investigated setup hindered the selectivity for short-chain PFASs (i.e., PFBuA, PFPeA, and PFBuS) retention as highlighted in paragraph 3.1, the weak anionic interactions seemed to be slightly suppressed using the adopted conditions. This can be observed by the close elution in both first and second dimension of a good number of analyte pairs with different ionic moieties were characterized by a close elution in both first and second dimension, such as PFPeA - Hx-FTS, PFBuS - PFHxA, Oc-FTS - PFPeS, D-FTS - PFHpS, and PFDA - PFNS. This finding could be also addressed to the steep gradient elution adopted on the C18 stationary phase. For this reason, a multi-step gradient program was

adopted in setup 3. **Fig. 2C** shows the obtained scatter plot of the two-dimensional space separation of the investigated PFAS standards. The implemented multi-step gradient allowed to gain retention and reduce peak width in the second dimension, also enhancing the contribution of the separation promoted by the ionic interactions deriving from the first dimension. In this regard, it should be underlined that the analyte pairs "co-eluting" in the setup 2, were more spaced in the 2D scatter plot. These findings are also supported by the peak capacity increase of the 30% ($n'_{c,2D}$ from 80 to 104). Additionally, by adopting setup 3, the retention time on the second dimension of PFBuA was slightly increased compared to the early elution obtained with setup 2 (**Fig. 2B**). The changes made in the second dimension gradient reduced the extent of the breakthrough of PFPeA and PFBUS (**Figs. S5B3** and **S5C3**), whereas a relevant peak splitting still occurred for PFBuA (**Fig. S5A3** and **S6G** of the *Supplementary Materials*). To further optimize the utilization of the 2DLC space we focused on the RPWAX method by setting the starting percentage of the 1D organic modifier at 35% (with a gradient time of 26 min) and adjusting the pH of the eluents of the first dimension at pH 4 to stabilize and increase the ionic interactions during the increase of the organic modifier, while reducing the hydrophobic ones, to obtain a better peak shape on the RPWAX column (see setup 4 in **Table 1**). Moreover, to increase the peak coverage along the second dimension, the first gradient ramp on the second dimension was set at 0.09 min.

In order to address the breakthrough and peak splitting, we investigated the use of Solid Phase Assisted Modulation (SPAM) and ASM (**Figure S4B**). The use of SPAM here should allow to (i) reduce the content of the organic fraction from the first dimension using a make-up eluent (i.e., second dimension eluent A), (ii) to concentrate the analytes between the two separation dimension (as long that these have sufficient retention on the trap columns [39]), and (iii) increase the 1D gradient flow rate, allowing to run 1D separations with higher gradient volumes without increasing the injection volumes in the 2D [40,41]. Due to the relative high percentage of the organic modifier through the entire first-dimension

chromatographic run, 1D eluents were diluted by a factor of ten by the make-up eluent during SPAM. The use of setup 4 in combination with SPAM brought considerable advantages. First, the width of the peaks in 1D has been reduced by approximately 50% (e.g., averaged width at peaks' baseline from 2 to 1 min), in favour of the increase of the predicted peak capacity of the system. **Fig. 2D** shows the scatter plot of the two-dimensional space obtained with setup 4, in which PFBA retention was significantly increased from a t_{R2D} <0.1 min with setup 3, to 0.35 min in setup 4. In addition, the spread in the second dimension of the analytes > C8 and different ionic groups (e.g., PFOSA, PFDS, and PFUnDA) has been increased. An even more relevant outcome deriving from the application of SPAM, was the peak splitting suppression, as shown in **Fig. S5A4**, and **S6H**. Finally, the predicted peak capacity of the system was increased from 104 to 181 using these conditions, as well as the $1/\beta$ was reduced to 0.5. Despite these advantages, the investigated setups still did not use completely the available separation space. In particular, relatively long times were needed to detect the first analytes (t0 around 10 min), the first dimension peak capacity was relatively low (27) and undersampling of peaks was present due to relatively long second dimension separation (1 min). We therefore tested two methods trying to address the limitation described for the first (setup 5) and second (setup 6) dimensions. Briefly, in setup 5 the flow rate was increased to 0.1 mL min^{-1} to deliver gradients with higher gradient volumes, whereas the initial percentage of the organic modifier was lowered to 25% and the gradient time was set to 50 min in order not to have a too rapid elution of the training set. The new method significantly reduced peak width in the first dimension (from 2 to 1 min), therefore the modulation time was set at 0.55 min to avoid under-sampling. Then, in setup 6 the gradient time in the second dimension was 0.4-fold increased relatively to the modulation time (see **Table 1**). **Figs. 2E** and **2F** show the scatter plots of the two-dimensional space obtained with setups 5 and 6, respectively. The conditions adopted in setup 6 increased the predicted peak capacity of about 60% compared to setup 4 ($n'_{c,2D}$=342), which was not significantly affected by the slight increase

of 1/β (**Table S4**). Furthermore, the increased chromatographic efficiency in the first and second dimension (**Fig. 2F**) allowed to (i) significantly separate short-chain PFASs (i.e., <C8) and to (ii) maintain the separation of PFOSA from the rest of the training set, highlighting selective elution areas of the two-dimensional space as a function of the polarity of the PFASs. In addition, a sufficient resolution between the homologous series of carboxylic and sulphonic acids is also appreciable. Considering these results, the conditions selected for the analysis of AFFF mixtures were the one adopted in setup 6.
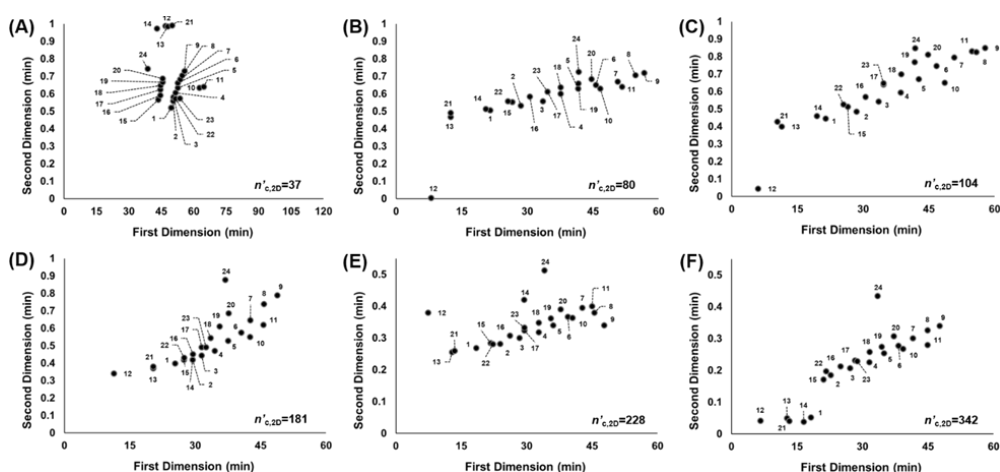
Figure 2 − Two-dimensional scatterplots of the setups investigated for the 2D-LC-HRMS method development, reported in Table 1. The predicted peak capacity ($n'_{c,2D}$) is reported in the bottom right of each scatter plot. (A) Setup 1, (B) setup 2, (C) setup 3, (D) setup 4, (E) setup 5, and (F) setup 6. 1=PFHxA, 2=PFHpA, 3=PFOA, 4=PFNA, 5=PFDA, 6=PFUnDA, 7=PFDoDA, 8=PFTrDA, 9=PFTeDA, 10=Me-FSA, 11=Et-FSA, 12=PFBuA, 13=PFPeA, 14=PFBuS, 15=PFPeS, 16=PFHxS, 17=PFHpS, 18=PFOS, 19=PFNS, 20=PFDS, 21=Hx-FTS, 22=Oc-FTS, 23=D-FTS, 24=PFOSA. Abbreviation meanings are reported in Table S1 of the Supplementary Materials.

7.3.3   Analysis of AFFF samples

The developed 2D-LC-HRMS method was applied to the analysis of two AFFF mixtures, i.e. 3M and ORC (see paragraph 2.2). Raw data are available at https://massive.ucsd.edu/ProteoSAFe (MSV000090462). **Table 2** reports the average number (n=3) of features detected with the developed method for an analysis conducted under equivalent conditions in 1D-LC, on both samples. In the

case of the 3M sample, the number of detected features was doubled compared to a one-dimensional approach (from 109 in 1D-LC to 211 in LC×LC), whereas for the ORC sample, which resulted to be "a less complex" sample using both approaches, the detected features were increased of about 0.5 times (from 85 in 1D-LC to 125 in LC×LC). These results show how the use of the developed two-dimensional approach led to a significant increase in the features detected in both mixtures by exploiting the two-dimensional chromatographic separation. The sensitivity of the non-targeted method (evaluated on the training set) was also improved, as reported by the LOI ranges in **Table 2**, showing a gain in terms of minimum concentration for the feature identification of about an order of magnitude lower, due to the analytes pre-concentration of the analytes during the transfer on the C18 cartridges with the SPAM configuration. Furthermore, the inter-day reproducibility of the non-targeted method (**Table 2**) was also evaluated through the stability of the retention time of the internal standards (i.e., M2PFOA and MPFOS) in individual and pooled samples, showing optimal values of percentage of relative standard deviation (0.01-0.04%). As commonly performed in non-targeted protocols, data quality was also explored by means of PCA performed on the intensities of the detected features after raw data alignment. PCA score plot of scaled and centered data on the first two dimensions (total explained variance of 68.30%) is reported in **Figure S7** of the *Supplementary Materials*, in which the quality of the data is highlighted by the proximity of pooled samples to the origin of the axes and reproducibility is confirmed by the good grouping of the injection replicates. Additionally, the difference in the (i) number of detected feature (**Table 2**) and (ii) their intensities in the two samples, also supported by the PCA score plot, further corroborate the future application of the developed method to the analysis of AFF-impacted environmental samples. **Figure 3A** and **3B** report the two-dimensional scatter plots of the features detected in the 3M and ORC samples, respectively, analyzed with the developed LC×LC-HRMS method.

Table 2 − Figures of merit of the non-targeted LC×LC-HRMS method: (i) average number of features detected in the investigated samples (n=3) and limits of identification (LOIs, µg L⁻¹) in comparison with equivalent 1D-LC methods and (ii) inter-day retention time (t_R) reproducibility (n=3) of internal standards in both individual and pooled samples, expressed as percentage of the relative standard deviation (RSD%).

Detected features			
	3M	ORC	LOI
1D-LC-HRMS	109	85	5-1000
LC×LC-HRMS	211	125	0.4-100
t_R RSD%			
	Samples	Pooled samples	
M2PFOA	0.04	0.02	
MPFOS	0.01	0.01	

Both figures depict a good separation in the two-dimensional space, confirming the reliability of the predicted peak capacity value of the system (**Fig. 2F**). Interestingly, three elution regions in the two-dimensional space were recognized, in accordance with those observed for the scatter plot of the training set of PFASs in **Fig. 2F**. In fact, the detected features eluted in distinct retention areas, the first one with t_{R1D} = 0-30 min and t_{R2D} = 0.0-0.1 min ascribable to the region of the most polar features, the second with t_{R1D} = 15-50 min and t_{R2D} = 0.1 -0.4 min where components with increasing hydrophobicity elute (e.g., chain elongation of homologue series), and finally the third with t_{R1D} = 0-45 min and t_{R2D} = 0.4-0.5 min in which features are potentially characterized by different ionic groups, but similar in terms of selectivity in the second dimension. **Figures 3C** and **3D** show the mass defect plots obtained using the -CF_2- scale mass difference, adopted to identify, among the detected features, the potential PFASs that occur inside the investigated samples, resulting in 39 and 15 potential PFAS features in 3M and ORC, respectively. Among them, it was possible to identify several exact masses (confirmed also by t_{R1D} and t_{R2D} match) of homologous series that also occur in the used training set of PFASs (see **Table S5**), i.e. PFASAs (248.94-598.93), PFACAs (268.98-318.98), and FTSs (426.97-526.96), which demonstrated the reliability of the developed LC×LC-HRMS method in PFAS separation and detection inside real samples.
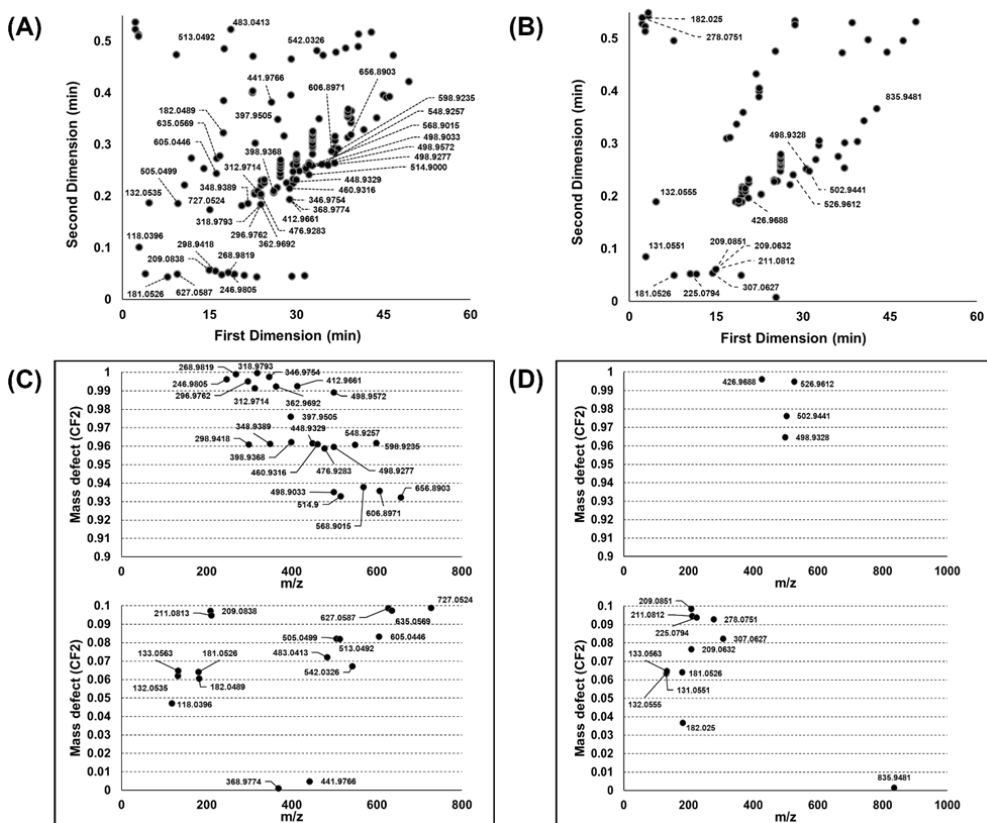
Figure 3 − Two-dimensional scatterplots of (A) 3M and (B) ORC samples obtained using the developed LC×LC-HRMS method. Labels refer to the matched exact masses of features identified as potential PFASs according to the KMD plots reported in figure (C) and (D), for 3M and ORC samples, respectively.

Furthermore, not only the number of detected 2D features was different between the two samples (**Table 2**), but also the composition in terms of potential PFASs identified. In detail, the 3M sample appears to be richer in compounds of interest than ORC, such as perfluorinated sulfonic salts and carboxylic acids, as already elsewhere reported [14,21]. Differently, the ORC sample seems to be characterized by the occurrence of polyfluorinated compounds, such as sulfonic telomer salts. Finally, the difference in the number of 2D features between LC×LC and 1D-LC is also reflected in the number of potential identified PFASs. In fact, the exploited selectivity in first and second dimension made detectable 17 and 5 additional *m/z* for 3M and ORC, respectively. Such increase in detectability is

given for example by the features at *m/z* 498.9277, 498.9572, and 498.9033, which were identified in the optimized LC×LC configuration in the 3M sample, while in 1D-LC only the first ion occurred, as also displayed in **Figure S8** of the *Supplementary Materials*.

7.4 Conclusions

2D-LC techniques offer numerous advantages in deciphering the chemical space of complex mixtures, increasing the peak capacity of chromatographic systems. In this research, it was possible to develop a comprehensive RPWAX×RPLC-HRMS technique for the separation of a training set of 24 PFAS, studying the selectivity of two different RPWAX stationary phases in combination with the traditional C18. In fact, through (i) the analysis of the orthogonality plots of 1D-LC and (ii) of the 2D-LC scatter plots, it was possible to optimize the chromatographic conditions (e.g., organic modifier, pH, flow rate, etc.) and select the best modulation strategy (i.e., SPAM) to obtain an optimal distribution in the two-dimensional space ($n'_{c,2D} \sim 300$) of the investigated analytes as a function of their polarity (i.e., ionic group and chain length). Subsequently, the LC×LC-HRMS method was applied to the characterization of AFFF mixtures, demonstrating the gain in terms of the number of detected features (up to two times, using only the chromatographic domain) with respect to 1D-LC. Furthermore, as observed for the training set, the same chromatographic differentiation was observed for the features recognized as potential PFASs according to KMD analysis, confirming the reliability of the developed method. Finally, it is important to underline that the LC×LC-HRMS method here presented paves the way for future 2D-LC applications in the analysis of poly- and perfluorinated compounds of concern in environmental matrices impacted by AFFF.

References

[1] B.W. Pirok, A.F. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J. Sep. Sci., 41 (2018) 68-98 https://doi.org/10.1002/jssc.201700863.

Section 7

[2] D.R. Stoll, P.W. Carr, Two-dimensional liquid chromatography: a state of the art tutorial, Anal. Chem., 89 (2017) 519-531 https://doi.org/10.1021/acs.analchem.6b03506.

[3] Y. Chen, L. Montero, O.J. Schmitz, Advance in on-line two-dimensional liquid chromatography modulation technology, TrAC, Trends Anal. Chem., 120 (2019) 115647 https://doi.org/10.1016/j.trac.2019.115647.

[4] P. Donato, F. Cacciola, P.Q. Tranchida, P. Dugo, L. Mondello, Mass spectrometry detection in comprehensive liquid chromatography: basic concepts, instrumental aspects, applications and trends, Mass Spectrom. Rev., 31 (2012) 523-559 https://doi.org/10.1002/mas.20353.

[5] N.L. Kuehnbaum, P. Britz-McKibbin, New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era, Chem. Rev., 113 (2013) 2437-2468 https://doi.org/10.1021/cr300484s.

[6] B.W. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent developments in two-dimensional liquid chromatography: fundamental improvements for practical applications, Anal. Chem., 91 (2018) 240-263 https://doi.org/10.1021/acs.analchem.8b04841.

[7] M. Grübner, A. Dunkel, F. Steiner, T. Hofmann, Systematic Evaluation of Liquid Chromatography (LC) Column Combinations for Application in Two-Dimensional LC Metabolomic Studies, Anal. Chem., 93 (2021) 12565-12573

[8] A.-K. Rausch, R. Brockmeyer, T. Schwerdtle, Development, validation, and application of a multi-method for the determination of mycotoxins, plant growth regulators, tropane alkaloids, and pesticides in cereals by two-dimensional liquid chromatography tandem mass spectrometry, Anal. Bioanal. Chem., 413 (2021) 3041-3054 https://doi.org/10.1021/acs.analchem.1c01857.

[9] P.F. Brandão, A.C. Duarte, R.M. Duarte, Comprehensive multidimensional liquid chromatography for advancing environmental and natural products research, TrAC, Trends Anal. Chem., 116 (2019) 186-197 https://doi.org/10.1016/j.trac.2019.05.016.

[10] V. Pérez-Fernández, L.M. Rocca, P. Tomai, S. Fanali, A. Gentili, Recent advancements and future trends in environmental analysis: Sample preparation, liquid chromatography and mass spectrometry, Anal. Chim. Acta, 983 (2017) 9-41 https://doi.org/10.1016/j.aca.2017.06.029.

[11] K. Horváth, J.N. Fairchild, G. Guiochon, Detection issues in two-dimensional on-line chromatography, J. Chromatogr. A, 1216 (2009) 7785-7792 https://doi.org/10.1016/j.chroma.2009.09.016.

[12] L. Ciofi, L. Renai, D. Rossini, C. Ancillotti, A. Falai, D. Fibbi, M.C. Bruzzoniti, J.J. Santana-Rodriguez, S. Orlandini, M. Del Bubba, Applicability of the direct injection liquid chromatographic tandem mass spectrometric analytical approach to the sub-ng L^{-} 1 determination of perfluoro-alkyl acids in waste, surface, ground and drinking water samples, Talanta, 176 (2018) 412-421 https://doi.org/10.1016/j.talanta.2017.08.052.

[13] K. Dasu, X. Xia, D. Siriwardena, T.P. Klupinski, B. Seay, Concentration profiles of per-and polyfluoroalkyl substances in major sources to the

environment, J. Environ. Man., 301 (2022) 113879 https://doi.org/10.1016/j.jenvman.2021.113879.

[14] B.J. Place, J.A. Field, Identification of novel fluorochemicals in aqueous film-forming foams used by the US military, Environ. Sci. Tech., 46 (2012) 7120-7127 https://doi.org/10.1021/es301465n.

[15] S. Peshoria, D. Nandini, R. Tanwar, R. Narang, Short-chain and long-chain fluorosurfactants in firefighting foam: a review, Environ. Chem. Lett., 18 (2020) 1277-1300 https://doi.org/10.1007/s10311-020-01015-8.

[16] T. Wang, Y. Wang, C. Liao, Y. Cai, G. Jiang, Perspectives on the inclusion of perfluorooctane sulfonate into the Stockholm convention on persistent organic pollutants, ACS Publications, 2009 https://doi.org/10.1021/es900464a.

[17] Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC (Text with EEA relevance), (2006).

[18] Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy Text with EEA relevance, (2013).

[19] Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption (recast) (Text with EEA relevance), (2020).

[20] R.A. García, A.C. Chiaia-Hernández, P.A. Lara-Martin, M. Loos, J. Hollender, K. Oetjen, C.P. Higgins, J.A. Field, Suspect screening of hydrocarbon surfactants in AFFFs and AFFF-contaminated groundwater by high-resolution mass spectrometry, Environ. Sci. Tech., 53 (2019) 8068-8077 https://doi.org/10.1021/acs.est.9b01895.

[21] L.A. D'Agostino, S.A. Mabury, Identification of novel fluorinated surfactants in aqueous film forming foams and commercial surfactant concentrates, Environ. Sci. Tech., 48 (2014) 121-129 https://doi.org/10.1021/es403729e.

[22] Y.-S. Luo, N.A. Aly, J. McCord, M.J. Strynar, W.A. Chiu, J.N. Dodds, E.S. Baker, I. Rusyn, Rapid Characterization of Emerging Per-and Polyfluoroalkyl Substances in Aqueous Film-Forming Foams Using Ion Mobility Spectrometry–Mass Spectrometry, Environ. Sci. Tech., 54 (2020) 15024-15034 https://doi.org/10.1021/acs.est.0c04798.

[23] W.J. Backe, T.C. Day, J.A. Field, Zwitterionic, cationic, and anionic fluorinated chemicals in aqueous film forming foam formulations and groundwater from US military bases by nonaqueous large-volume injection HPLC-MS/MS, Environ. Sci. Tech., 47 (2013) 5226-5234 https://doi.org/10.1021/es3034999.

[24] J. Janda, K. Nödler, H.-J. Brauch, C. Zwiener, F.T. Lange, Robust trace analysis of polar (C2-C8) perfluorinated carboxylic acids by liquid

chromatography-tandem mass spectrometry: method development and application to surface water, groundwater and drinking water, Environ. Sci. and Poll. Res., 26 (2019) 7326-7336 https://doi.org/10.1007/s11356-018-1731-x.

[25] R. Montes, R. Rodil, L. Placer, J.M. Wilms, R. Cela, J.B. Quintana, Applicability of mixed-mode chromatography for the simultaneous analysis of C1-C18 perfluoroalkylated substances, Anal. Bioanal. Chem., 412 (2020) 4849-4856 https://doi.org/10.1007/s00216-020-02434-w.

[26] X.-Z. Niu, L. Abrell, R. Sierra-Alvarez, J.A. Field, J. Chorover, Analysis of hydrophilic per-and polyfluorinated sulfonates including trifluoromethanesulfonate using solid phase extraction and mixed-mode liquid chromatography-tandem mass spectrometry, J. Chromatogr. A, (2022) 462817 https://doi.org/10.1016/j.chroma.2022.462817.

[27] L. Zhang, Q. Dai, X. Qiao, C. Yu, X. Qin, H. Yan, Mixed-mode chromatographic stationary phases: recent advancements and its applications for high-performance liquid chromatography, TrAC, Trends Anal. Chem., 82 (2016) 143-163 https://doi.org/10.1016/j.trac.2016.05.011.

[28] X. Cai, Z. Guo, X. Xue, J. Xu, X. Zhang, X. Liang, Two-dimensional liquid chromatography separation of peptides using reversed-phase/weak cation-exchange mixed-mode column in first dimension, J. Chromatogr. A, 1228 (2012) 242-249 https://doi.org/10.1016/j.chroma.2011.06.042.

[29] F. Li, X. Su, S. Bäurer, M. Lämmerhofer, Multiple heart-cutting mixed-mode chromatography-reversed-phase 2D-liquid chromatography method for separation and mass spectrometric characterization of synthetic oligonucleotides, J. Chromatogr. A, 1625 (2020) 461338 https://doi.org/10.1016/j.chroma.2011.06.042.

[30] S. Samanipour, J.W. O'Brien, M.J. Reid, K.V. Thomas, Self adjusting algorithm for the nontargeted feature detection of high resolution mass spectrometry coupled with liquid chromatography profile data, Anal. Chem., 91 (2019) 10800-10807 https://doi.org/10.1021/acs.analchem.9b02422.

[31] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, A cross-platform toolkit for mass spectrometry and proteomics, Nat. Biotechnol., 30 (2012) 918-920 https://doi.org/10.1038/nbt.2377.

[32] X. Li, D.R. Stoll, P.W. Carr, Equation for peak capacity estimation in two-dimensional liquid chromatography, Anal. Chem., 81 (2009) 845-850 https://doi.org/10.1021/ac801772u.

[33] K.A. Barzen-Hanson, S.C. Roberts, S. Choyke, K. Oetjen, A. McAlees, N. Riddell, R. McCrindle, P.L. Ferguson, C.P. Higgins, J.A. Field, Discovery of 40 classes of per-and polyfluoroalkyl substances in historical aqueous film-forming foams (AFFFs) and AFFF-impacted groundwater, Environ. Sci. Tech., 51 (2017) 2047-2057 https://doi.org/10.1021/acs.est.6b05843.

[34] A.L. Myers, K.J. Jobst, S.A. Mabury, E.J. Reiner, Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products, J. Mass Spectrom., 49 (2014) 291-296 https://doi.org/10.1002/jms.3340.

[35] B. González-Gaya, N. Lopez-Herguedas, A. Santamaria, F. Mijangos, N. Etxebarria, M. Olivares, A. Prieto, O. Zuloaga, Suspect screening workflow comparison for the analysis of organic xenobiotics in environmental water samples, Chemosphere, 274 (2021) 129964 https://doi.org/10.1016/j.chemosphere.2021.129964.

[36] T.H. Walter, M. Blaze MT, C. Boissel, Electrospray ionization mass spectrometry ion suppression/enhancement caused by column bleed for three mixed-mode reversed-phase/anion-exchange high-performance liquid chromatography columns, Rapid Commun. Mass Spectrom., 35 (2021) e9098 https://doi.org/10.1002/rcm.9098.

[37] J. Hammer, J.-H. Haftka, P. Scherpenisse, J. Hermens, P. De Voogt, Investigating hydrophilic and electrostatic properties of surfactants using retention on two mixed-mode liquid chromatographic columns, J. Chromatogr. A, 1571 (2018) 185-192 https://doi.org/10.1016/j.chroma.2018.08.024.

[38] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active solvent modulation: a valve-based approach to improve separation compatibility in two-dimensional liquid chromatography, Anal. Chem., 89 (2017) 9260-9267 https://doi.org/10.1021/acs.analchem.7b02046.

[39] M.J. den Uijl, T. Roeland, T.S. Bos, P.J. Schoenmakers, M.R. van Bommel, B.W. Pirok, Assessing the Feasibility of Stationary-Phase-Assisted Modulation for Two-Dimensional Liquid-Chromatography Separations, J. Chromatogr. A, (2022) 463388 https://doi.org/10.1016/j.chroma.2022.463388.

[40] R.J. Vonk, A.F. Gargano, E. Davydova, H.L. Dekker, S. Eeltink, L.J. De Koning, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography with stationary-phase-assisted modulation coupled to high-resolution mass spectrometry applied to proteome analysis of Saccharomyces cerevisiae, Anal. Chem., 87 (2015) 5387-5394 https://doi.org/10.1021/acs.analchem.5b00708.

[41] A. Baglai, M.H. Blokland, H.G. Mol, A.F. Gargano, S. van der Wal, P.J. Schoenmakers, Enhancing detectability of anabolic-steroid residues in bovine urine by actively modulated online comprehensive two-dimensional liquid chromatography–high-resolution mass spectrometry, Anal. Chim. Acta, 1013 (2018) 87-97 https://doi.org/10.1016/j.aca.2017.12.043.

8 Innovative combination of thermal desorption with on-line solid phase extraction-reversed phase liquid chromatography applied to targeted nutrimetabolomics in human biofluids.

Abstract

This study proposed for the first time the use of thermal desorption in the on-line solid phase extraction coupled with re-versed phase liquid chromatography (on-line SPE-RP-LC) using mixed mode polymeric sorbents. This analytical strategy was applied to the on-line SPE-RP-LC targeted analysis of a model set of 34 human gut metabolites characterized by heter-ogeneous physicochemical properties (e.g., octanol–water partition coefficient in the range -0.3 – 3.4). The novel tempera-ture-assisted on-line SPE approach was investigated in comparison with the conventional desorption strategies at ambient temperature based on the use of (i) an optimized elution gradient or (ii) an organic desorption followed by post-column dilution. The thermal desorption strategy has been shown to be better performing and suitable for the development of a reliable and sensitive method for the analysis of the model group of analytes in urine and serum. In more detail, under the optimized experimental conditions (Isolute ENV+ sorbent, desorption temperature 120°C, desorption time 5.5 min), the proposed method provided negligible matrix effects in both biofluids for almost all target analytes. Moreover, method quantification limits were in the ranges 0.026-7.2 µg L^{-1} and 0.033-23 µg L^{-1} for urine and serum, respectively, i.e., comparable to or lower than those reported in methods previously published.

8.1 Introduction

On-line solid phase extraction (on-line SPE) coupled with liquid chromatography (LC) is a high analytical throughput technique, suitable to an effective sample preconcentration and matrix removal in the quantification of a wide range of analytes within different contexts (e.g., environmental and biological fluid analysis), both on untreated and pre-treated samples [1]. This technique is commonly applied in combination with triple quadrupole tandem mass spectrometry (QqQ) for quantitative analysis, due to the overall superior performances of this latter detection mode com-pared to UV or other mass analyzers [2]. The development of an on-line SPE-LC-QqQ method is highly complex due to the need to select sorbent phases suitable for retaining target analytes and to make analytes desorption compatible with subsequent chromatographic separation. When the investigated molecules contained strong hydro-phobic moieties, sorbent phases based on silica silanized with alkyl (e.g., C18) [3] or aromatic (e.g., phenyl) [4] groups in combination with reversed phase (RP) chromatography produced satisfactory results, guaranteeing a good analytes retention without effects of breakthrough, nor analyte over-retention. However, if target analytes cover a wide range of physicochemical properties and one or more polar moieties are present in the molecule, the use of mixed-mode sorbents (i.e., providing multimodal interactions such as π-π, dipole-dipole, hydrogen bond, and hydrophobic interactions) is recommended, since they offer a very high retentive capacity towards a broad spectrum of target compounds [5]. Due to this, mixed-mode sorbents require in most cases high percentages of organic solvent for their detachment, thus making this desorption strategy poorly compatible with the retention of polar analytes on RP chromatographic stationary phases. On the other hand, the use of eluents suitable for RP-LC involves an inadequate detachment of the analytes from the sorbent phase, with poor focus in the analytical column and/or presence of carryover effects [6,7]. In order to maintain an RP-LC compatible with the on-line SPE, an alternative desorption strategy relies on the organic desorption followed by post-column dilution (OD-PCD), which consists

in the elution of the cartridge with a low volume (tens of µL) of organic solvent and the post-cartridge dilution with an aqueous solvent, as similar as possible to the initial chromatographic conditions [8-9]. The post-cartridge dilution allows to significantly improve the head retention of polar analytes, which otherwise would not be retained or underwent to chromatographic distortions (e.g., peak broadening and splitting). An alternative strategy for using on-line mixed-mode SPE is its combination with hydrophilic interaction liquid chromatography (HILIC), which is compatible with the use of high percentages of organic solvent at the beginning of the elution gradient [10-11]. However, HILIC is characterized by complex retention mechanisms and a poor reproducibility of retention times with even small variations of mobile phase modifiers (e.g. buffer strength and pH), thus resulting in a kind of "on–off chromatography" for some com-pounds and unpredictable chromatographic behaviours compared to RP [12]. On the other hand, during the desorption step the elution strength of the aqueous mobile phase can be improved by increasing the temperature of the eluent, which weakens the non-covalent bonds responsible for analyte retention and increases the elution strength of water, due to the lowering of its permittivity [13].

Accordingly, in this study, thermal desorption (TD) was investigated for the first time in on-line SPE-RP-LC-QqQ, using mixed-mode sorbents and water as SPE eluent. This de-sorption strategy was tested on a model group of structurally heterogeneous analytes comprising 34 human gut nutrimetabolites characterized by (i) a wide range of physico-chemical properties, (ii) different physiological conjugations (i.e., glucuronides and sulfonates), and (iii) structural isomerism (i.e., isobars), thus representing a challenging analytical issue. In detail, the chromatographic separation of the model analytes was investigated using different RP stationary phases and eluent combinations and finally optimized focusing on the resolution of critical isobaric groups. Then, TD was studied in comparison with (i) OD-PCD and (ii) conventional desorption (i.e. using the optimized elution gradient without heating), demonstrating the much better performance of this novel desorption approach in terms of analyte focusing in the analytical column

and sensitivity. Finally, the on-line TD-SPE-RP-LC-QqQ method was evaluated for matrix effect and recovery in serum and urine, as relevant applications in nutrimetabolomics [14].

8.2 Materials and methods

8.2.1 Chemicals and reagents

LC-MS grade water, methanol and acetonitrile were obtained from Carlo Erba (Milan, Italy). Ultrapure water (resistivity > 18 MΩ) was obtained from a Milli-Q system (Millipore, Billerica, MA, USA). Formic acid (HCOOH, MS grade, purity ≥ 99.0%) and trifluoroacetic acid (CF3COOH, MS grade, purity ≥ 99.0%) were purchased from Sigma-Aldrich (St. Louis, MO, USA) and Merck (Darmstadt, Ger-many), respectively. The chromatographic columns used in this study were the endcapped pellicular Kinetex® C18, biphenyl (BP), and phenyl-hexyl (P-H) columns (Phenomenex, Torrance, CA, USA), 100 × 3 mm, particle size 2.6 μm, each. The SPE cartridges Isolute ENV+ and Oasis HLB used in this study were obtained from Biotage (Uppsala, Sweden) and Waters (Milford, MA, USA), respectively. These sorbents were selected based on information regarding their thermal stability (up to 160°C), given by the manufacturers. Full details of the two cartridges are reported in **Table S1** of the *Supporting Information*.

8.2.2 Target analytes and preparation of stock and working solutions

The target compounds of this study (TRC, Ontario, Cana-da) belonged to the following metabolites categories: hydroxybenzenes, benzoic acids, cinnamic acids, urolithins, stilbenoids, and hippuric acids. Their full name, abbreviation, and physicochemical properties are reported in **Table S2** of the *Supporting Information*. Luteolin (LUT) and sinapic acid (SIN), used as internal standards, were purchased from Extrasynthese (Genay, France). Standard stock solutions of individual compounds were prepared by accurate weighing and dissolution in LC-MS grade methanol to give a concentration of 1 g L^{-1}, apart from EA, for which a 0.1 g L^{-1} standard stock solution was prepared due to its lower solubility. A multicompound standard solution at a concentration of 10 mg L^{-1} of each

compound was prepared by proper dilution of the standard stock solutions in LC-MS grade methanol. The latter was further diluted with LC-MS water up to (i) 250 µg L^{-1} for direct injection (DI) analysis, and (ii) 25 µg L^{-1} for the optimization of the on-line SPE method, as operative standard solutions. The standard stock solutions and the 10 mg L^{-1} solution were stored in the dark at -20 °C, whereas the operative solutions were stored in the dark at $+4$ °C.

8.2.3   *Instrumental configurations tested*

Instrumental analysis was performed on a Shimadzu (Kyoto, Japan) chromatographic system coupled with a 5500 QTrap mass spectrometer (Sciex, Framingham, MA, USA), equipped with a Turbo V® interface by an ESI probe (see Section S3 of the *Supporting Information* for source de-pendent parameters). Source and compound dependent parameters were optimized as respectively described in section S3 and **Table S3** of the *Supporting Information*. QqQ analysis was carried out using Multiple Reaction Monitoring (MRM) in negative ionization mode. The optimization of the on-line SPE extraction process on mixed-mode sorbents was carried out by investigating three different desorption approaches (see **Figure S1 A-C** of the *Supporting Information*), i.e. (A) conventional desorption by exposing the cartridge to the optimized chromatographic gradient, without any heating, (B) OD-PCD, and (C) TD. All these on-line SPE-LC configurations used a low pressure gradient quaternary pump Nexera X2 LC-30AD (pump 1) and one isocratic pump LC-20AD XR, devoted to the on-line SPE procedure (pump 2). An additional pump (pump 3) is used in the OD-PCD configuration to manage a make-up solvent (i.e., acidified LC-MS grade water), according to the configuration reported in the **Fig. S1B**. A CTO/20AC thermostatted column compartment equipped with the analytical column, a SIL-30AC auto-injector equipped with a 2-mL sample loop, a DGU-20A 5R degassing unit and a CBM-20A module controller were also used. On-line SPE was performed using a two-position six-port switching valve (model HT, Vici, Schenkon, Switzerland) housing the sorbent cartridge. In the TD

configuration (**Fig. S1C**), two experimental devices, schematically illustrated in **Figure 1**, were tested.
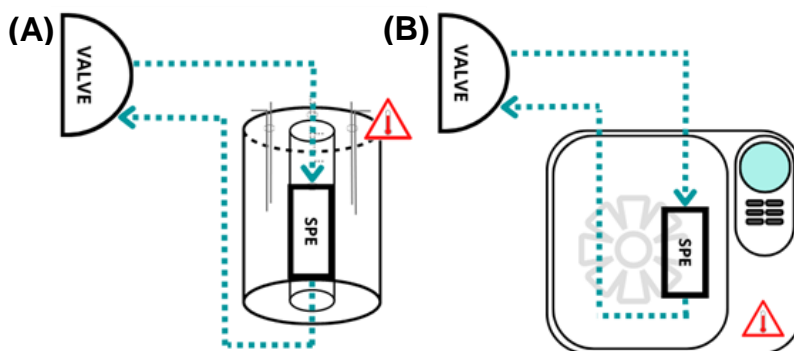
Figure 1 – Schematic illustration of the thermal desorption systems 1 (A) and 2 (B).

8.2.4   Evaluation of the on-line SPE configurations

Preliminarily, to test the applicability of the TD configuration, it was investigated the occurrence of degradation phenomena of target analytes due to their exposure at high temperatures. In particular, these tests were performed in triplicate by placing vials containing the independent aliquots of operative standard solutions (see section 2.2) in a stainless-steel holder and heating at temperatures of 90 and 120 °C, for a total time of 5.5 min (i.e., simulating the heating procedure of the SPE cartridge as described in section 2.5). The evaluation of the thermal degradation of the 34 metabolites was carried out by comparing the mean chromatographic areas (n = 3) of each analyte in DI (A_{DI}) vs the values obtained after the controlled exposure at the aforementioned temperature levels (A_{SPE}), using 250 μg L^{-1} operative solution. The three on-line SPE configurations (**Fig. S1**) were investigated keeping constant the optimized chromatographic conditions. Desorption temperatures of 60, 90, and 120 °C and methanol volumes of 50 and 100 μL were tested for TD and OD-PCD configurations, respectively. The performances of the extraction and desorption procedures were evaluated in terms of (i) peak shape and, when possible, (ii) extraction efficiency (EE%,

Equation 1) on all the 34 human gut metabolites, using the aforementioned operative standard solutions at 25 µg L^{-1}.

$$EE\% = \left(\frac{A_{SPE}}{A_{DI}} \cdot 100 \right) \qquad (1)$$

A_{SPE} and A_{DI} are the mean chromatographic areas (n = 3) obtained under on-line SPE configuration and DI, respectively. Within day and between day precision associated to EE% was also assessed. For each human gut metabolite, instrumental detection (iLODs) and quantitation limits (iLOQs) were determined by applying the best performing on-line SPE-TD-LC-MS/MS method (i.e. the TD system 2 set at 120°C) on operative standard solutions. To this aim, **Equation 2** was used, where σ_b is the standard deviation of the procedural blank and S is the slope of the calibration curve in ultrapure water [15].

$$iLOD \ (iLOQ) = \frac{3.3 \ (10) \cdot \sigma_b}{S} \qquad (2)$$

The instrumental linearity range was also determined, using iLOQ as the lower calibration level.

8.2.5   Sample preparation

Serum and urine samples were collected within a randomized, two-arm, acute intervention trial on ten volunteers, who were orally administered a single dose of bilberry or blueberry (i.e., *Vaccinium myrtillus* and *Vaccinium corymbosum*) supplement. Before analysis, serum and urine aliquots were thawed and pre-treated as described elsewhere [16]. For the evaluation of the figure of merits of the analytical method and for the quantitative analysis, pooled samples of serum and urine were prepared separately using the biofluids of all the volunteers, regardless of the treatment administered.

8.2.6   Figures of merit of the best performing on-line SPE-TD-LC-MS/MS method on biofluids

The performances of the method were evaluated in real samples in terms of percentages of matrix effect (ME%), apparent recovery (AR%), recovery (R%),

as well as method detection (MDLs), and quantification limits (MQLs) [15]. These assessments were performed on a pool of on-line SPE eluate fractions from the loading of urine and serum pool aliquots, recovered before entering the analytical column and corresponding to peak-free retention time intervals [6].

ME% was determined using **Equation 3**:

$$ME\% = \left(\frac{S_{matrix}}{S_{solvent}} \cdot 100 \right) - 100 \qquad (3)$$

where S_{matrix} and $S_{solvent}$ are the slopes of calibration lines of each metabolite and internal standards, in matrix and in LC-MS water, respectively. AR% values of each metabolite and internal standards were evaluated on the aforementioned eluate spiked at two concentration levels (i.e., 25 and 50 μg L^{-1}), analyzed as described below (section 2.8), and calculated according to **Equation 4**, where A_{matrix} is the chromatographic area in spiked eluate pool and $A_{solvent}$ is the chromatographic area in LC-MS water. These samples were analyzed from run-to-run over 3 days to estimate a "between days" precision.

$$AR\% = \left(\frac{A_{matrix}}{A_{solvent}} \cdot 100 \right) \qquad (4)$$

R % was defined using **Equation 5**:

$$R\% = AR\% - ME\% \qquad (5)$$

MDLs and MQLs were assessed according to **Equation 6**:

$$MDL\ (MQL) = \frac{iLOD\ (iLOQ)}{AR\%} \qquad (6)$$

8.2.7 *Operative on-line SPE-LC-MS/MS conditions*

In the operative conditions, the TD system 2 was adopted, performing the sample analysis as follows. The 2-mL loop was filled with 500 μL standard solution/sample at 10 μL s^{-1} (filling time of the loop equal to 50 s). Meanwhile, the six-port switching valve was set on the loading position, allowing the analytical column to be conditioned by pump 1 with acidified water (0.1% HCOOH) and 2% of acetonitrile, at a flow rate of 0.7 mL min^{-1}. When the loop

was filled, the elution program of pump 1 started. Simultaneously, pump 2 delivered a solution of ultrapure water acidified at pH=2 with HCl at 1.0 mL min^{-1} for 2.8 min, thus loading the 500 µL of the standard solution/sample onto the Isolute ENV+ cartridge. During the loading phase, the cartridge was kept at 25° C. Subsequently, the temperature of the oven that houses the cartridge was brought to 120 °C in about 20 s and maintained for 5.5 min before switching the valve to the injection position, and back-flushing the SPE cartridge by pump 1 for the entire gradient program, thus allowing the transfer of target analytes to the analytical column. The chromatographic conditions are summarized hereafter. Chromatographic column: Kinetex P-H; eluent "A": acidic LC-MS grade water (0.1% HCOOH); eluent "B": acidic methanol (0.1% HCOOH); eluent "C": acidic acetonitrile (0.1% HCOOH); flow rate: 0.70 mL min^{-1}; temperature: 40 °C; elution gradient: eluent "C" 2.0% for the entire duration of the analysis; eluent "B" percentage 2.0 % for 1 min followed by its increase at 2%/min for 13 min, 5%/min for 7.5 min, and final 100% of "B" in 3 min. The system is then brought to the initial conditions in 0.5 min. for re-equilibrating the analytical column.

8.2.8 *Data analysis*

The chromatographic optimization as well as the on-line SPE method development were conducted using Analyst (version 1.6.2) and Multiquant (version 3.0.2) softwares (Sciex).

8.3 **Results and discussions**

8.3.1 *Chromatographic optimization*

Three superficially porous stationary phases (i.e., Kinetex C18, BP, and P-H), characterized by the same geometry, were tested (section 2.1). Chromatographic optimization was performed by evaluating the setups reported in **Table S4**, using the 250 µg L^{-1} operative solution and an injection volume equal to 50 µL (i.e., injecting 12.5 ng of each analyte). The acidity of mobile phases was kept constant at pH = 2.7±0.1 with 0.1% formic acid, to promote the neutral speciation of the target analytes (see **Table S2**), without affecting their electrospray ionization. In

this latter regard, when stronger acidity was used (e.g., 1% formic acid or 0.1% trifluoroacetic acid), a relevant drop of sensitivity (data not shown) was observed as elsewhere reported in literature [17]. The target set of analytes included five isobaric groups: 2,3-DBA/2,5-DBA, TRES/UR-A, iFA-SULF/FA-SULF, QUE-3-GLU/QUE-7-GLU, and UR-A-GLU/iUR-A-GLU/TRES-GLU. Since the first three groups showed chromatographic resolutions (R_{Chrom}, **Equation S1**) > 1.5 in all the investigated setups, the discussion of the optimization procedure focused only on UR-A-GLU/iUR-A-GLU/TRES-GLU and QUE-3-GLU/QUE-7-GLU groups. **Table S4** also reports the R_{Chrom} values of the critical isobaric groups for each investigated experimental condition. Setups 1-2 (i.e., C18 combined with methanol as organic eluent) were not able to resolve none of the isobaric groups, while the replacement of C18 with BP increased the retention of all isobaric groups and particularly of TRES-GLU, thus providing its resolution from the urolithin derivatives (R_{Chrom} = 2.4). The use of P-H gave rise to results very similar to those achieved with C18, i.e. without any baseline separation of the isobaric groups (data not shown). The organic strength of mobile phase was increased using acetonitrile instead of methanol, evaluating also different flow rates (from 0.4 to 0.6 mL min^{-1}), thus achieving lower analytes retention and a general improvement of R_{Chrom}. These setups (i.e., 4-to-6 in **Table S4**) effectively allowed for obtaining baseline separations of the two urolithins and between iUR-A-GLU and RES-GLU. In addition, an increase in the separation of the quercetin derivatives was observed, being the best results obtained at 0.5 mL min^{-1} (setup 5, R_{Chrom} = 0.6). These findings could be addressed to the multiple selectivity mechanisms exerted by the BP stationary phase (e.g., shape, polarizability, and polarity of the target analytes) [18]. In an attempt to improve the separation of pairs only partially resolved in setup 5, the P-H was also tested. In fact, based on literature [19], this stationary phase should have a greater capacity to adopt different spatial orientations of the aromatic ring, thus potentially giving rise to a wider interaction capacity of the stationary phase with the target analytes, compared to BP column. The results of configurations 7 and 8 clearly showed the

complete loss of resolution of urolithin derivatives, which were however well separated from resveratrol, while quercetin derivatives remained only partially resolved. Further tests on the P-H column focused on the use of methanol and methanol/acetonitrile eluents. As illustrated in **Table S4**, the use of a constant 2% of acetonitrile within a gradient water/methanol as described in Section 2.5, provided a good compromise of the R_{Chrom} values, leading to a satisfying resolution for all critical isobaric groups, when the flow rate of 0.7 mL min^{-1} was adopted (see also **Figure S2** of the Supporting Information). This finding is justified by the combination of the organic strength of methanol towards analytes and the π-π disrupting properties of acetonitrile towards P-H, the latter increasing the partition rate to the mobile phase [20]. Further increments of flow rate, probably helpful to improve critical resolutions, were not allowed due to an incomplete spray formation in the ion source. Thus, setup 11 was confirmed as the operative chromatographic method to be coupled with on-line SPE.

8.3.2 Development of the on-line SPE-LC-MS/MS method

Once the optimal chromatographic conditions were identified and the system configured for on-line SPE, the three different desorption strategies described in **Fig. S1** were investigated to maximize the EE% of target metabolites. In detail, on-line SPE was studied performing (i) the conventional desorption, i.e. using the optimized LC elution gradient keeping the sorbent at ambient temperature (**Fig. S1 A**), (ii) the OD-PCD (**Fig. S1 B**), and (iii) the TD using the same optimized gradient elution adopted for the conventional desorption (**Fig. S1 C**), investigating three tempera-ture levels (i.e., 60, 90, and 120 °C). Furthermore, the two commercially available mixed-mode cartridges Isolute ENV+ and Oasis HLB, fully described in **Table S1**, were evaluated comparatively. Under the experimental conditions adopted for loading the sample onto the on-line cartridge (i.e., ultrapure water acid-ified at pH=2 with HCl at flow rate of 1.0 mL min^{-1} for 2.8 min), breakthrough phenomena were never highlighted in both sorbents, thus evidencing the strong retention capacity of the mixed-mode phases tested here against the target analytes.

8.3.2.1   Conventional desorption

The development of the on-line SPE method was initially investigated by adopting the configuration reported in **Fig. S1 A**, i.e., exposing the SPE cartridge to the flow of the mobile phase from pump 1, for the entire chromatographic run. Desorption performances of both cartridges were studied in these experimental conditions for all target analytes, highlighting very similar behaviours of the two stationary phases. In detail, the increase in the organic strength along the elution gradient was not enough rapid to promote the prompt detachment of target analytes from the stationary phase and their focusing, thus generating broad and unresolved elution profiles, particularly for more polar analytes. As an example, **Figure S3** of the *Supporting Information* illustrates the desorption from the Isolute ENV+ cartridge of four representative analytes (i.e., CA-GLU, 5-FQA, 2,3-DBA, and UR-A). These compounds were chosen to cover almost the entire chromatographic window and a wide range of values of the logarithm of octanol-water partition constants (Log D) at the loading and desorption pH values (pH=2-3, see **Table S2** of the *Supporting Information*). These findings are in agreement with results elsewhere re-ported for the desorption from mixed-mode sorbents of gut metabolites similar to the ones here investigated [21,22]. In fact, in these papers, the analytes detachment required the use of 100% organic solvents, while the chromatographic gradient adopted here reached 100% organic in about 26 min, thus determining the aforementioned distortion of the chromatographic profile and making unfeasible any evaluation of the analytes EE% using these on-line SPE conditions.

8.3.2.2   Organic desorption with post-cartridge dilution

Based on literature information [8-9], to solve the issue of over-retention of analytes on on-line SPE sorbents, the OD-PCD configuration illustrated in **Fig. S1 B** was also tested on both Isolute ENV+ and Oasis HLB cartridges. Using this configuration, the rapid detachment of the analytes from the sorbent phase can be promoted by introducing a small volume of organic solvent into the cartridge.

Then, the analytes can be focused at the top of the analytical column by diluting the cartridge eluate with an aqueous solution. Despite the apparent simplicity of this approach, a high complexity is experienced, since it is essential to optimize the volumes of both desorption organic solvent and focusing aqueous solution, as well as the speed and timing with which they are dispensed. Tests were conducted for this purpose, trying to identify the best compromise among an efficient desorption of the analytes, their focusing and the sensitivity of the on-line SPE method, which is of course affected by the post-cartridge dilution. As a general consideration, better recoveries were obtained using the Isolute ENV+ cartridge. The best results were obtained by configuring pump 1 to deliver 100 µL (i.e., approximately the void volume of the cartridge) of methanol acidified with 0.1% formic acid (i.e., 0.14 min at 0.7 mL min^{-1}) in the cartridge, immediately after the loading step. At the end of this dosage, the cartridge was exposed for 1 min to the starting chromatographic conditions (i.e., 2% acetonitrile, 2% methanol, and 96% water) while the post-cartridge dilution was performed by dosing 100 µL of 0.1% aqueous formic acid (make-up solvent) with pump 3 (i.e., 1 min at 0.1 mL min^{-1}). Afterwards, the valve was rotated to the loading position (**Fig. S1 B**, left side) and the elution gradient started as described in section 2.8. Compared to the previous configuration, the analytes de-sorption and focusing was significantly improved, even though the peak shapes were still broad in comparison with direct injection (see **Figure S4** of the *Supporting Information*). Moreover, the chromatographic resolution of UR-A-GLU/iUR-A-GLU was drastically worsened and QUE-3-GLU/QUE-7-GLU co-eluted completely (see **Figure S5 A-B** of the *Supporting Information*). In spite of these drawbacks, the optimized OD-PCD approach allowed for calculating the recovery of all target metabolites for both cartridges. **Figure 2** illustrates the individual recoveries obtained for the four representative metabolites previously selected (**Fig. 2 A**) and the mean recovery of all target analytes (**Fig. 2 B**), highlighting the better performances of the Isolute ENV+ cartridge.

Figure 2 – Extraction efficiency of selected analytes (box A) and all analytes (mean value, box B) from the OASIS HLB and Isolute ENV+ sorbents, using the organic desorption with post-cartridge dilution. Error bars represent standard deviation.

The lower recovery obtained with the Oasis HLB sorbent should be attributed to an insufficient desorption capacity of the eluent used, rather than a lower retention capacity of the sorbent during the loading phase, as no breakthrough was observed, as mentioned above. A lower recovery of Oasis HLB compared to Isolute ENV+ was elsewhere re-ported for the off-line SPE of various phenolic compounds in aqueous matrices, although no clear interpretation of this result was given [23]. In detail, EE% values achieved with the Oasis HLB and the Isolute ENV+ was 26±15 (range: 2-56) and 40±23 (range: 10-106), respectively. Therefore, even under the best OD-PCD conditions, i.e. with the ENV+ cartridge, the EE% was unsatisfactory, with three analytes (i.e., VA, TRES, and TRES-SULF) showing values < 20% and those related to further eleven compounds included in the range 20-30%.

8.3.2.3   *Thermal desorption*

Due to the problems encountered with the OD-PCD strategy, it was decided to adopt an innovative SPE approach compared to what has been reported in the literature so far, investigating the effect of temperature on the partition be-tween the sorbent phase and the eluent, during the desorption step from the cartridge. Using a desorption strategy based on the increase in the cartridge temperature, the detachment of analytes from the sorbent should be facilitated due to the (i) weakening of the non-covalent bonds between the analyte and the sorbent phase

and (ii) decrease in the permittivity of water, which brings its eluent strength closer to that of an alcohol at room temperature [24-25]. The thermal degradability tests carried out at 90 and 120 °C on the 34 target metabolites highlighted in all cases a good thermal stability (see **Figure S6** of the *Supporting Information*), thus allowing to test the heating-assisted desorption strategy for the whole model group of analytes. The TD system 1 (**Fig. 1 A**) was the first to be tested showing better desorption with the ENV+ than the Oasis HLB, similarly to what was observed with the OD-PCD approach. **Figure 3** shows the results obtained for the ENV+ cartridge with the TD system 1 set at the temperatures of 60°C and 90°C. These results refer to the recovery of the aforementioned four representative gut metabolites (**Fig. 3 A**) and to the mean recovery of the whole group of target analytes (**Fig. 3 B**).
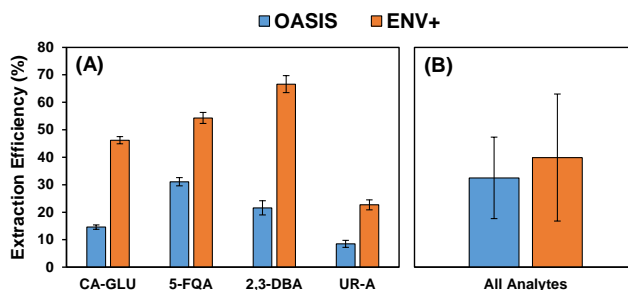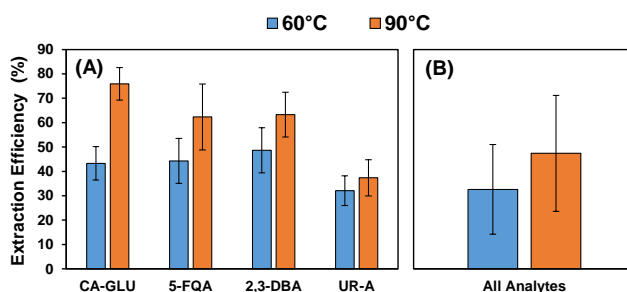

Figure 3 – Extraction efficiency of selected analytes (box A) and all analytes (mean value, box B) from the Isolute ENV+ sorbent, using the thermal desorption system 1 set at 60 and 90°C. Error bars represent standard deviation.

Moreover, **Figure S7 A-D** of the *Supporting Information* illustrates for the same four individual analytes the peak shape obtained under the best experimental conditions, i.e. 90°C. The heating of the cartridge improved the detachment of the analytes from the cartridge, whose peak shape appears much more homogeneous and narrower, compared to the profile observed for the OD-PCD technique, even though the resolutions of the critical isobaric pairs UR-A-GLU/iUR-A-GLU and QUE-3-GLU/QUE-7-GLU were partially or totally lost (see **Fig. S5 C-D**). The recovery was also improved, with a clear effect of the temperature increase from 60°C to 90°C, the latter corresponding to a mean recovery value for all analytes

of about 50%. No further significant increase in the desorption temperature (i.e., up to 120-130 °C) was unfortunately possible with the TD system 1, because the target temperature was reached too slowly to allow a prompt detachment of the analytes, probably due to the characteristics of the cylindrical heaters used for the implementation of the device. To overcome this drawback the TD system 2 (**Fig. 1 B**) was tested. Based on the results achieved with the TD system 1, the temperatures of 90 °C and 120 °C were investigated on both sorbents also in the TD system 2. A prompt detachment of analytes and narrow desorption peaks were obtained also with this desorption system (**Figure S8** of the *Supporting Information*). Furthermore, similarly to findings obtained with OD-PCD and TD system 1, higher recoveries were achieved with the ENV+ than the OASIS HLB (**Figure 4**). The results obtained as a function of the temperature applied to the cartridge confirmed the trend previously observed on the TD system 1. In detail, at the desorption temperature of 120°C, the mean extraction efficiency was 95±25, with only two analytes characterized by EE% val-ues less than 60%, i.e. QUE-3-SULF and FA-SULF, both exhibiting values of about 30% (see **Table S5** of the *Supporting Information*).
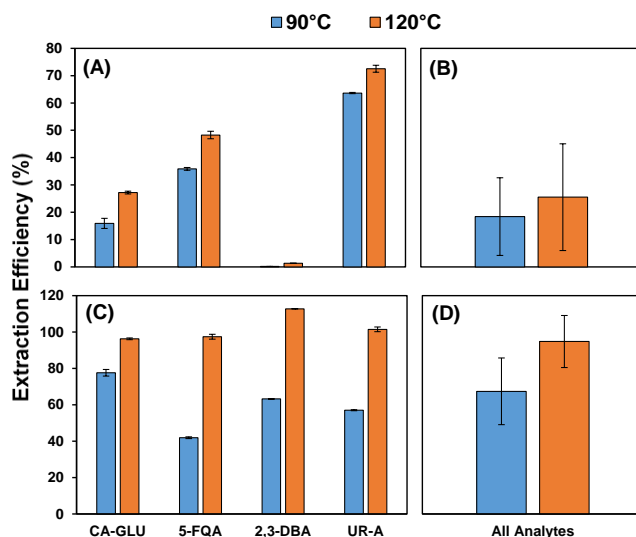
Figure 4 – Extraction efficiency of selected analytes (boxes A and C) and all analytes (mean value, boxes B and D) from OASIS HLB (boxes A and B) and Isolute ENV+ (boxes C and D) sorbents, using the ther-mal desorp-tion system 2 set at 90 and 120°C. Error bars represent standard deviation.

Even applying this desorption strategy, the critical resolutions of the isobaric pairs UR-A-GLU/iUR-A-GLU and QUE-3-GLU/QUE-7-GLU were not maintained (see **Figure S5 E-F**), thus highlighting this behaviour as a general drawback of the translation of the direct injection method into the online SPE one. The on-line SPE-LC-MS/MS chromatogram of the operative standard solution under the selected experimental conditions is reported in **Figure S9** of the *Supporting Information*. The experimental conditions based on the use of the ENV+ cartridge and the TD system 2 heated at 120°C were considered as optimal and therefore tested for the evaluation of the instrumental performances.

The vast majority of metabolites showed iLODs and iLOQs of tens-hundreds of ng L^{-1}, with the sole exceptions of VA, EA, and 2,3-HYBA, which exhibited sensitivities of few µg L^{-1} (**Table S5**). The linearity range was investigated from the iLOQ values up to 500 µg L^{-1}, showing good R^2 values (0.9962-0.9998) for the entire set of metabolites (**Table S5**).

8.3.3 *Method validation in urine and serum samples*

The on-line SPE method using the thermal desorption strategy 2 on the ENV+ cartridge was validated for the analysis of target gut metabolites and internal standards in serum and urine samples from an intervention study on bilberry and blueberry intake [16]. The validation involved the parameters mentioned in section 2.7, i.e., ME%, MDLs, MQLs, AR%, and R% (see **Table S6** of the *Supporting Information*). **Figure S10** of the *Supporting Information* illustrates the chromatographic profile of some analytes, obtained under the proposed experimental conditions in spiked ultrapure water (**Fig. S10 A-C**), urine (**Fig. S10 D-F**), and serum (**Fig. S10 G-I**).

8.3.3.1   Urine samples

Urine samples showed signal suppression rather than an increase in most cases. However, the effects due to the alteration of ionization efficiency in the ESI source was mostly negligible, being $|ME| \leq 20\%$ in almost all cases. The only significant exceptions were TRES and TRES-SULF with a suppressive effect of 26%, and above all SAL, which exhibited a signal suppression of 38%. AR% values, evaluated on target metabolites and internal standards at 25 µg L^{-1} and 50 µg L^{-1} (levels I and II in **Table S6**, respectively), were characterized by a high precision, being the relative standard deviation (RSD) associated with independent replicates (n = 3) in all cases ≤8%. Furthermore, the values of AR% determined for each compound at the two spiking levels were very similar and therefore a mean AR% was calculated and used to assess the values of R%. Overall, R% values ranged from 31% to 118%. In more detail, for 26 out of 35 analytes (note that for co-eluting QUE-3-GLU and QUE-7-GLU, a cumulative recovery was calculated) R% was included in the pseudo-quantitative range of 80-120%, while 7 compounds were recovered between 60% and 79%, and the remaining 2 in the range 30-59%. The method sensitivity was found similar to the instrumen-tal one, due to the limited effect of the matrix on both on-line SPE and ionization efficiency in the ESI source.

Accordingly, MDLs and MQLs were mostly included between tens and hundreds of ng/L, with the very few above-mentioned exceptions of VA, EA, and 2,3-DBA (MQLs of 2.0, 7.2, and 4.9 µg L^{-1}, respectively, see **Table S6**), to which QUE-3-SULF is added (MQL = 1.3 µg L^{-1}), due to its low value of AR%. Hence, method sensitivity here achieved is optimal for the quantification of target analytes in biological fluids, where their concentrations commonly range from nM to µM levels [26]. It is interesting to compare the performances of the on-line SPE method here optimized with SPE protocols elsewhere reported that use mixed mode sorbents (i.e., hydroxylated styrene–divinylbenzene copolymer and styrene–divinylbenzene copolymer functionalized with N-vinylpyrrolidone). One paper [21] focusing on off-line SPE shares with our study FA, VA, HVA, CA, 3-

HPAA, UR-A, UR-B, and SIN, for which only R% values were however reported. The comparison highlighted better performances of our method, with the only exception of HVA. Furthermore, the gain in analysis time obtained with the on-line method was significant, since the analysis of one sample took less than 30 minutes against about three hours of the off-line one.

8.3.3.2   *Serum samples*

Unlike urine, serum samples were mostly characterized by a signal enhancement. However, the effect of the matrix on compound ionization was in almost all cases negligible (i.e., |ME| ≤ 20%), being α-HHA and CA-GLU the only exceptions, with values of matrix effect of about 28% and 33%, respectively (**Table S6**). This finding is particularly interesting in light of the high complexity of the serum matrix and the relative simplicity of the proposed on-line protocol. Similar to the results shown in urine, the precision associated with the three independent replicates of the AR% assessment was also very high in serum, the RSD% values being ≤ 6% (**Table S6**). Overall, values of R% ranged from 13% to 113%. In more detail, only 10 out of 35 analytes showed R% ≥ 70%, whereas 11 analytes were recovered between 40% and 69% and the remaining 15 below 40%. These findings affected also the overall method sensitivity for serum analysis, evidencing in some cases MQLs in the 1-2 µg L^{-1} range (i.e., 3-MC, 2,5-DBA, 3-HPAA, and QUE-3-SULF) and even higher, such as VA (4.7 µg L^{-1}), EA (23.1 µg L^{-1}), and 2,3-DBA (20.4 µg L^{-1}). However, the other analytes exhibited comparable or slightly higher MQLs compared to urine, while remaining in the tens to hundreds ng L^{-1} range. Two papers have been previously published with a focus on the analysis of nutrimetabolites in serum using off-line SPE protocols based on mixed mode copolymer sorbents and therefore comparable with our study [21,27]. **Figure 5** illustrates by mean of boxplots the comparison of the values of ME%, R%, and MQL for the 8 (Ref. 19) and 23 (Ref. 24) compounds that are in common with this study. The proposed method generally involved a signal enhancement instead of the suppressive trend highlighted elsewhere (**Fig. 5 A**).
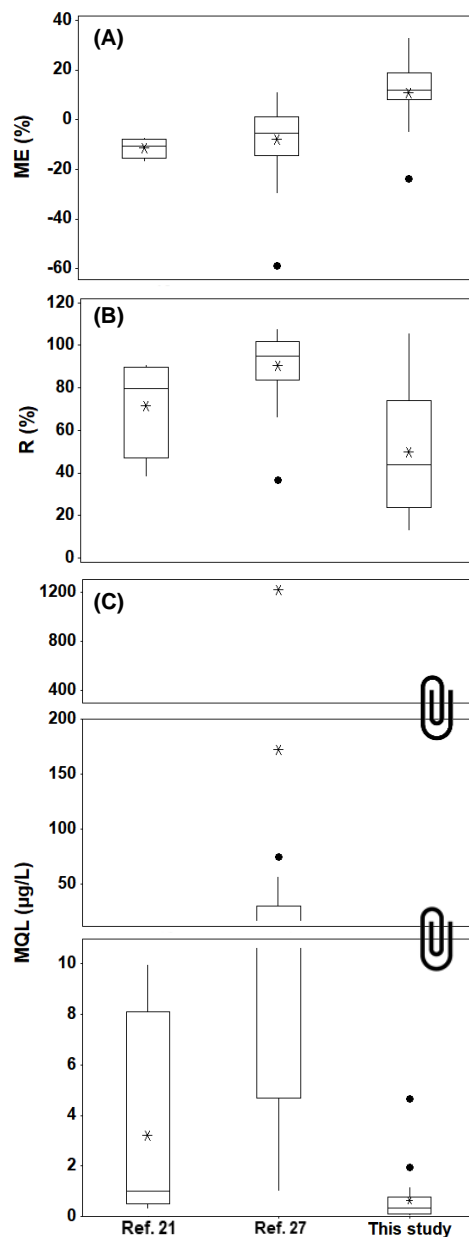
Figure 5 – Boxplot of the values of ME%, R% and MQL determined in Ref. 19 and 24, in comparison to this study. The boxplots refer respectively to 8 (Ref. 21) and 23 (Ref. 27) compounds in common with this study. The boxes rep-resent the interquartile range, while the line is the median; lower and upper whiskers extend to the maximum or min-imum data point within 1.5 box height from the top or bot-tom of the box; asterisk is the mean; black round symbols are the outliers.

However, the matrix effect values were mostly within the range ±20%, which is commonly considered negligible [13], thus evidencing the effectiveness of this kind of sorbents in the removal of the matrix constituents. The trend of R% (**Fig. 5 B**) found in this study was characterized by a greater dispersion and a lower overall average, especially in comparison with the method pro-posed by González-Domínguez and co-workers [24], which is more robust from this point of view. On the other hand, despite the higher recoveries reported elsewhere, MQLs found in our study (**Fig. 5C**) were much less dispersed and far lower than those achieved by the off-line SPE approaches under consideration. This result should be attributed to the straightforward procedure here proposed by using on-line SPE on 500 µL of sample aliquots, rather than to higher instrumental performance, as the sensitivity of the mass spectrometer used in this study is comparable to that of the instruments used in the other researches.

8.4   Conclusions

For the first time, temperature-assisted desorption was investigated for the detachment of analytes strongly retained on mixed mode polymer sorbents, demonstrating the feasibility of this analytical strategy in the on-line SPE approach and on highly complex biological matrices such as urine and serum. In this regard, it should be noted that after method optimization the ENV+ cartridge was used for about 150 injections showing the same chromatographic profile and extraction performances. This research is also the first on-line SPE study on a large group of nutrimetabolites with very different physicochemical properties. The optimized method is suitable for the RP-LC analysis of gut metabolites in targeted metabolomics platforms.

References

[1] Y. Liu, Q. Song, J. Zheng, J. Li, Y. Zhao, C. Li, Y. Song, P. Tu, Sensitive profiling of phenols, bile acids, sterols, and eicosanoids in mammalian urine by large volume direct injection-online solid phase extraction-ultra high performance liquid chromatography-polarity switching tandem mass spectrometry. RSC advances, 6 (2016) 81826-81837 https://doi.org/10.1039/C6RA13272C.

[2] J. G. van der Gugten, Tandem mass spectrometry in the clinical laboratory: a tutorial overview. Clinical Mass Spectrometry, 15 (2020) 36-43 https://doi.org/10.1016/j.clinms.2019.09.002.

[3] L. Ciofi, C. Ancillotti, U. Chiuminatto, D. Fibbi, B. Pasquini, M.C. Bruzzoniti, L. Rivoira, M. Del Bubba, Fully automated on-line solid phase extraction coupled to liquid chromatography–tandem mass spectrometry for the simultaneous analysis of alkylphenol polyethoxylates and their carboxylic and phenolic metabolites in wastewater samples, Analytical and bioanalytical chemistry, 408 (2016) 3331-3347 https://doi.org/10.1007/s00216-016-9403-5.

[4] D. Bury, V. N. Belov, Y. Qi, H. Hayen, D. A. Volmer, T. Brüning, H. M. Koch, Determination of urinary metabolites of the emerg-ing UV filter octocrylene by online-SPE-LC-MS/MS. Analytical Chemistry, 90 (2018) 944-951 https://doi.org/10.1021/acs.analchem.7b03996.

[5] D. Sýkora, P. Řezanka, K. Záruba, V. Král, Recent advances in mixed-mode chromatographic stationary phases, Journal of Separation Sciences, 42 (2019), 89-129 https://doi.org/10.1002/jssc.201801048.

[6] D. Rossini, L. Ciofi, C. Ancillotti, L. Checchini, M.C. Bruzzoniti, L. Rivoira, D. Fibbi, S. Orlandini, M. Del Bubba, Innovative combination of QuEChERS extraction with on-line solid-phase extract purification and pre-concentration, followed by liquid chromatography-tandem mass spectrometry for the determination of non-steroidal anti-inflammatory drugs and their metabolites in sewage sludge, Analytica chimica acta, 935 (2016) 269-281 https://doi.org/10.1016/j.aca.2016.06.023.

[7] T. Vega-Morales, Z. Sosa-Ferrera, J. Santana-Rodríguez, Development and optimisation of an on-line solid phase extraction coupled to ultra-high-performance liquid chromatography–tandem mass spectrometry methodology for the simultaneous determination of endocrine disrupting compounds in wastewater samples. Journal of Chromatography A, 1230 (2012) 66-76 https://doi.org/10.1016/j.chroma.2012.01.077.

[8] S. Huntscha, H. P. Singer, C. S. McArdell, C. E. Frank, J. Hollender, Multiresidue analysis of 88 polar organic micropollutants in ground, surface and wastewater using online mixed-bed multilayer solid-phase extraction coupled to high performance liquid chromatography–tandem mass spectrometry, Journal of Chromatography A, 1268 (2012) 74-83 https://doi.org/10.1016/j.chroma.2012.10.032.

[9] X. Ye, Z. Kuklenyik, L. L. Needham, A. M. Calafat, Automated on-line column-switching hplc-ms/ms method with peak focusing for the determination of nine environmental phenols in urine. Anal. Chem., 77 (2005) 5407-5413 https://doi.org/10.1021/ac050390d.

[10] N. Fontanals, R. Marcé, F. Borrull, New materials in sorptive extraction techniques for polar compounds, Journal of Chromatography A, 1152 (2007) 14-31 https://doi.org/10.1016/j.chroma.2006.11.077.

[11] J. Yan, J. Ding, G. Jin, D. Yu, L. Yu, Z. Long, Z. Guo, W. Chai, X. Liang, Profiling of Sialylated Oligosaccharides in Mammalian Milk Using Online Solid Phase Extraction-Hydrophilic Interaction Chromatography Coupled with

Negative-Ion Electrospray Mass Spectrometry, Analytical Chemistry, 90 (2018) 3174-3182. https://doi.org/10.1021/acs.analchem.7b04468.

[12] D. Zahn, I. J. Neuwald, T. P. Knepper, Analysis of mobile chemicals in the aquatic environment—current capabilities, limitations and future perspectives. Analytical and bioanalytical chemistry, 412 (2020) 4763-4784 https://doi.org/10.1007/s00216-020-02520-z.

[13] R. M. Smith, Extractions with superheated water, Journal of Chromatography A, 975 (2002) 31-46 https://doi.org/10.1016/S0021-9673(02)01225-6.

[14] M. Ulaszewska, M. Garcia-Aloy, N. Vázquez-Manjarrez, M. T. Soria-Florido, R. Llorach, F. Mattivi, C. Manach, Food intake bi-omarkers for berries and grapes. Genes & Nutrition, 15 (2020) 1-35 https://doi.org/10.1186/s12263-020-00675-z.

[15] C.V.A. Scordo, L. Checchini, L. Renai, S. Orlandini, M.C. Bruzzoniti, D. Fibbi, L. Mandi, N. Ouazzani, M. Del Bubba, Optimization and validation of a method based on QuEChERS extraction and liquid chromatographic–tandem mass spectrometric analysis for the determination of perfluoroalkyl acids in strawberry and olive fruits, as model crops with different matrix characteristics, Journal of Chromatography A, 1621 (2020) 461038 https://doi.org/10.1016/j.chroma.2020.461038.

[16] C. Ancillotti, M. Ulaszewska, F. Mattivi, M. Del Bubba, Untargeted metabolomics analytical strategy based on liquid chromatography/electrospray ionization linear ion trap quadrupole/orbitrap mass spectrometry for discovering new polyphenol metabolites in human biofluids after acute ingestion of Vaccinium myrtillus berry supplement, Journal of American Society of Mass Spectrometry, 30 (2019) 381-402 https://doi.org/10.1007/s13361-018-2111-y.

[17] L. Renai, C.V.A. Scordo, A. El Ghadraoui, S. Santana-Viera, J.J.S. Rodriguez, S. Orlandini, S. Furlanetto, D. Fibbi, D. Lambropoulou, M. Del Bubba, Quality by design optimization of a liquid chromatographic-tandem mass spectrometric method for the simultaneous analysis of structurally heterogeneous pharmaceutical compounds and its application to the rapid screening in wastewater and surface water samples by large volume direct injection, Journal of Chromatography A, 1649 (2021) 462225 https://doi.org/10.1016/j.chroma.2021.462225.

[18] M. Powell, M.B. D'Arcy, Liquid-phase separation of structurally-similar steroids using phenyl stationary phases, Journal of Liquid Chromatography & Related Technologies, (2013) https://doi.org/10.1080/10826076.2013.778635.

[19] C. Brown, L. Dornan, M. Muldoon, R. T. Hembre, P. Stevenson, P. Manesiotis, Comparison of three stationary phases in the separa-tion of polyphenyls by liquid chromatography. Journal of Chromatography A, 1671 (2022) 462992 https://doi.org/10.1016/j.chroma.2022.462992.

[20] S. González-Rubio, A. Ballesteros-Gómez, D. Carreras, G. Muñoz, S. Rubio, A comprehensive study on the performance of different retention mechanisms in sport drug testing by liquid chromatography tandem mass spectrometry, Journal of Chromatography B, 1178 (2021) 122821 https://doi.org/10.1016/j.jchromb.2021.122821.

[21] Gasperotti, M.; Masuero, D.; Guella, G.; Mattivi, F.; Vrhovsek, U. Development of a targeted method for twenty-three metabolites related to polyphenol gut microbial metabolism in biological samples, using SPE and UHPLC–ESI-MS/MS, Talanta, 128 (2014) 221-230 https://doi.org/10.1016/j.talanta.2014.04.058.

[22] C. Li, G. Ai, Y. Wang, Q. Lu, C. Luo, L. Tan, G. Lin, Y. Liu, Y. Li, H. Zeng, Oxyberberine, a novel gut microbiota-mediated metabolite of berberine, possesses superior anti-colitis effect: impact on intestinal epithelial barrier, gut microbiota profile and TLR4-MyD88-NF-κB pathway, Pharmacological Research, 152 (2020) 104603 https://doi.org/10.1016/j.phrs.2019.104603.

[23] D. Bratkowska, N. Fontanals, F. Borrull, P. Cormack, D. Sherrington, R. Marcé, Hydrophilic hypercrosslinked polymeric sorbents for the solid-phase extraction of polar contaminants from water, Journal of Chromatography A, 1217 (2010) 3238-3243 https://doi.org/10.1016/j.chroma.2009.08.091.

[24] A. Catenaccio, Y. Daruich, C. Magallanes, Temperature dependence of the permittivity of water, Chemical Physics Letters, 367 (2003) 669-671 https://doi.org/10.1016/S0009-2614(02)01735-9.

[25] Y. Daruich, C. Magallanes, L. Giordan, E. Garis, A. Catenaccio, Temperature dependence of the permittivity of some pure alcohols. Molecular Physics, 99 (2001) 77-79 https://doi.org/10.1080/00268970010000467.

[26] A. Rodriguez-Mateos, R.D. Pino-García, T.W. George, A. Vidal-Diez, C. Heiss, J.P. Spencer, Impact of processing on the bioavailability and vascular effects of blueberry (poly) phenols, Molecular Nutrition & Food Research, 58 (2014) 1952-1961 https://doi.org/10.1002/mnfr.201400231.

[27] R. González-Domínguez, O. Jáuregui, P. Mena, K. Hanhineva, F.J. Tinahones, D. Angelino, C. Andrés-Lacueva, Quantifying the human diet in the crosstalk between nutrition and health by multi-targeted metabolomics of food and microbiota-derived metabolites, International Journal of Obesity, 44 (2020) 2372-2381 https://doi.org/10.1038/s41366-020-0628-1.

8.5 Supplementary materials

S1 Sorbent phases for on-line solid-phase-extraction

Table S1 – Details of the on-line SPE cartridge used in this study.

SPE cartridge	Geometry (mm)	Particle size (μm)	Chemistry	Pore-size (Å)	Surface Area (m²/g)	Thermal stability
Isolute ENV+	30 x 2.1	40	Hyper crosslinked hydroxylated styrene-divinylbenzene copolymer	800	1000	<160° C
Oasis HLB	20 x 2.1	25	Hyper crosslinked divinylbenzene – vinyl pyrrolidone copolymer	80	N/A	≤160° C

N/A: not available.

S2 Target (poly)phenol metabolites

Table S2 – List of the model group of gut metabolites, CAS numbers, their abbreviations and selected physicochemical properties taken from the Chemicalize cheminformatics platform: molecular weight (MW, Da), pKa (most acidic, T=25 °C), partitioning constants at loading and desorption pH; N/A = not available. The asterisk in bracket indicates the compounds chosen as representative of the whole model group of analytes and for which the extraction efficiency and peak shape under the different desorption conditions is shown.

Compound Name	CAS	Abbreviation	MW	pKa	Log D	
					pH=2.0	pH=3.0
Ferulic acid	1135-24-6	FA	194	3.97	1.67	1.63
Ferulic acid-4-O-glucuronide	86321-24-6	FA-GLU	370	3.05	-0.31	-0.57
Ferulic acid 4-O-sulfate	86321-29-1	FA-SULF	274	-2.25	-1.19	-1.30
Isoferulic acid 3-O-sulfate	1258842-19-1	iFA-SULF	274	-2.2	-1.19	-1.35
Vanillic acid	121-34-6	VA	168	4.16	1.17	1.14
Homovanillic acid	306-08-1	HVA	182	3.74	1.14	1.08
5-O-Feruloylquinic acid (*)	40242-06-6	5-FQA	368	3.33	-0.14	-0.29
Salicylic acid	69-72-7	SA	138	2.79	1.91	1.56
Caffeic acid	331-39-5	CA	180	3.34	1.51	1.37
Caffeic acid 4-O-glucuronide (*)	1093679-71-0	CA-GLU	356	2.95	-0.46	-0.77
Dihydro caffeic acid 3-O-sulfate	1187945-70-5	DCA-SULF	262	-2.23	-0.75	-0.78
Chlorogenic acid	327-97-9	CHA	354	3.33	-0.29	-0.43
Ellagic acid	476-66-4	EA	302	5.54	2.32	2.31
2,3-dihydroxybenzoic acid (*)	303-38-8	2,3-DBA	154	2.56	1.57	1.10
2,5-dihydroxybenzoic acid	490-79-9	2,5-DBA	154	2.53	1.56	1.08
3,4-Dihydroxybenzoic acid-3-O-sulfate	76496-11-2	3,4-DBA-SULF	234	-2.70	-1.18	-1.21
Hippuric acid	495-69-2	HA	179	3.59	0.51	0.43
α-hydroxyhippuric acid	16555-77-4	α-HHA	195	2.97	0.19	-0.08
3-hydroxyphenylacetic acid	621-37-4	3-HPAA	152	4.07	1.30	1.27
Abscisic acid	14375-45-2	AA	264	4.74	2.09	2.09
3-Methoxycathecol	934-00-9	3-MC	140	9.56	1.21	1.21
Quercetin-3-O-glucuronide	22688-79-5	QUE-3-GLU	478	2.65	-0.18	-0.61
Quercetin 3-O-sulfate	60889-05-6	QUE-3-SULF	382	-2.50	-0.96	-0.97
Quercetin 7-O-glucuronide	38934-20-2	QUE-7-GLU	478	2.72	-0.14	-0.53
Trans-resveratrol	501-36-0	TRES	228	8.49	3.40	3.40
Trans-resveratrol 3-O-glucuronide	387372-17-0	TRES-GLU	404	3.14	1.42	1.22
Trans-resveratrol 3-O-sulfate	553662-69-4	TRES-SULF	308	-2.13	0.56	0.55
Urolithin A (*)	1143-70-0	UR-A	228	8.41	2.50	2.50
Urolithin B	1139-83-9	UR-B	212	8.87	2.81	2.81
Urolithin C	165393-06-6	UR-C	244	8.16	2.20	2.20
Urolithin D	131086-98-1	UR-D	260	8.18	2.54	2.54
Urolithin A-3-O-glucuronide	1365982-52-0	UR-A-GLU	404	3.19	0.53	0.34
Urolithin B-3-O-glucoronide	823806-74-2	UR-B-GLU	388	3.19	0.83	0.64
Isourolithin A-3-O-glucuronide	N/A	iUR-A-GLU	404	N/A	N/A	N/A

S3 On-line SPE-LC-MS/MS analysis

Source dependent parameters were set as the following: Curtain gas (CUR) 40, Collision Activated Dissociation gas (CAD) medium, Temperature (TEM) 750 °C, Gas 1 (GS1) 60, Gas 2 (GS 2) 40, IonSpray Voltage (IS) -4500 V.

Table S3 – Optimized MS/MS parameters of target analytes. CE collision energy (V); DP declustering potential (V); EP entrance potential (V); CXP collision cell exit potential (V); precursor and product ions mass are in Da.

| Compound	Q1	Q3	DP	EP	CE	CXP	Compound	Q1	Q3	DP	EP	CE	CXP
FA 1	193	134	-50	-10	-20	-12	UR-C 1	243	187	-190	-7	-40	-10
FA 2	193	149	-50	-10	-15	-14	UR-C 2	243	171	-190	-7	-36	-12
VA 1	167	108	-120	-12	-28	-14	UR-C 3	243	199	-190	-7	-34	-16
VA 2	167	91	-120	-12	-29	-14	UR-B-GLU 1	387	211	-80	-5	-32	-17
VA 3	167	152	-120	-12	-20	-15	UR-B-GLU 2	387	113	-80	-5	-23	-17
3-MC 1	139	124	-60	-13	-20	-10	UR-B-GLU 3	387	227	-80	-5	-44	-44
3-MC 2	139	123	-60	-13	-32	-10	UR-A-GLU 1	403	227	-80	-10	-36	-13
3-MC 3	139	95	-60	-13	-42	-10	UR-A-GLU 2	403	113	-80	-10	-19	-13
SA 1	137	93	-45	-7	-25	-10	UR-A-GLU 3	403	300	-80	-10	-30	-10
SA 2	137	65	-45	-7	-43	-20	iUR-A-GLU 1	403	227	-87	-9	-30	-10
CA 1	179	135	-65	-10	-25	-12	iUR-A-GLU 2	403	113	-87	-9	-23	-11
CA 2	179	134	-65	-10	-35	-10	CA-GLU 1	355	135	-80	-9	-18	-10
CHA 1	353	191	-170	-13	-25	-20	CA-GLU 2	355	179	-80	-9	-20	-15
CHA 2	353	85	-170	-13	-60	-10	CA-GLU 3	355	227	-80	-9	-33	-11
EA 1	301	145	-180	-10	-50	-15	3,4-DBA-SULF 1	233	109	-80	-11	-33	-16
EA 2	301	229	-180	-10	-35	-20	3,4-DBA-SULF 2	233	153	-80	-11	-23	-11
EA 3	301	284	-180	-10	-40	-20	QUE-3-SULF 1	381	301	-64	-5	-18	-10
2,3-DBA 1	153	109	-80	-8	-22	-12	QUE-3-SULF 2	381	151	-64	-5	-40	-20
2,3-DBA 2	153	108	-80	-8	-34	-12	QUE-3-SULF 3	381	179	-64	-5	-40	-9
2,3-DBA 3	153	91	-80	-8	-36	-12	TRES-GLU 1	403	113	-106	-4	-27	-19
2,5-DBA 1	153	108	-80	-8	-32	-10	TRES-GLU 2	403	227	-106	-4	-28	-14
2,5-DBA 2	153	81	-80	-8	-22	-10	TRES-GLU 3	403	185	-106	-4	-52	-12
2,5-DBA 3	153	53	-80	-8	-27	-10	TRES-SULF 1	307	227	-196	-12	-27	-25
α-HHA 1	194	73	-62	-8	-20	-9	TRES-SULF 2	307	185	-196	-12	-42	-19
α-HHA 2	194	120	-62	-8	-20	-9	TRES-SULF 3	307	143	-196	-12	-49	-11
3-HPAA 1	151	65	-58	-8	-27	-10	UR-D 1	259	213	-185	-9	-40	-24
3-HPAA 2	151	107	-58	-11	-11	-10	UR-D 2	259	242	-185	-9	-36	-11
3-HPAA 3	151	106	-58	-11	-38	-10	UR-D 3	259	185	-185	-9	-50	-15
HA 1	178	77	-68	-9	-20	-10	FA-SULF 1	273	134	-80	-7	-38	-16
HA 2	178	132	-68	-9	-22	-10	FA-SULF 2	273	178	-80	-7	-33	-17
HA 3	178	102	-68	-9	-32	-10	FA-SULF 3	273	149	-80	-7	-28	-14
AA 1	263	153	-150	-7	-14	-18	5-FQA 1	367	193	-70	-8	-24	-19
AA 2	263	203	-150	-7	-39	-20	5-FQA 2	367	134	-70	-8	-46	-19
AA 3	263	122	-150	-7	-46	-18	5-FQA 3	367	149	-70	-8	-28	-19
TRES 1	227	185	-156	-7	-25	-14	iFA-SULF 1	273	193	-114	-9	-20	-15
TRES 2	227	143	-156	-7	-34	-14	iFA-SULF 2	273	134	-114	-9	-40	-24
TRES 3	227	117	-156	-7	-41	-14	iFA-SULF 3	273	178	-114	-9	-31	-25
UR-A 1	227	226	-80	-10	-34	-25	QUE-7-GLU 1	477	301	-90	-10	-38	-25
UR-A 2	227	198	-80	-10	-46	-25	QUE-7-GLU 2	477	151	-90	-10	-50	-25
HVA 1	181	122	-107	-7	-19	-12	QUE-7-GLU 3	477	179	-90	-10	-44	-25
HVA 2	181	121	-107	-7	-35	-12	DCA-SULF 1	261	181	-167	-8	-26	-7
HVA 3	181	77	-107	-7	-27	-12	DCA-SULF 2	261	137	-167	-8	-31	-13
FA-GLU 1	369	178	-178	-6	-22	-16	DCA-SULF 3	261	109	-167	-8	-38	-9
FA-GLU 2	369	134	-178	-6	-48	-16	SIN 1	223	149	-20	-8	-27	-15
FA-GLU 3	369	193	-178	-6	-36	-16	SIN 2	223	164	-20	-8	-18	-20
QUE-3-GLU 1	477	301	-80	-10	-30	-25	LUT 1	285	133	-40	-10	-45	-15
QUE-3-GLU 2	477	151	-80	-10	-48	-25	LUT 2	285	151	-40	-10	-35	-15
QUE-3-GLU 3	477	179	-80	-10	-43	-25							
UR-B 1	211	167	-160	-7	-36	-23							
UR-B 2	211	139	-160	-7	-40	-15							
UR-B 3	211	169	-160	-7	-36	-21							
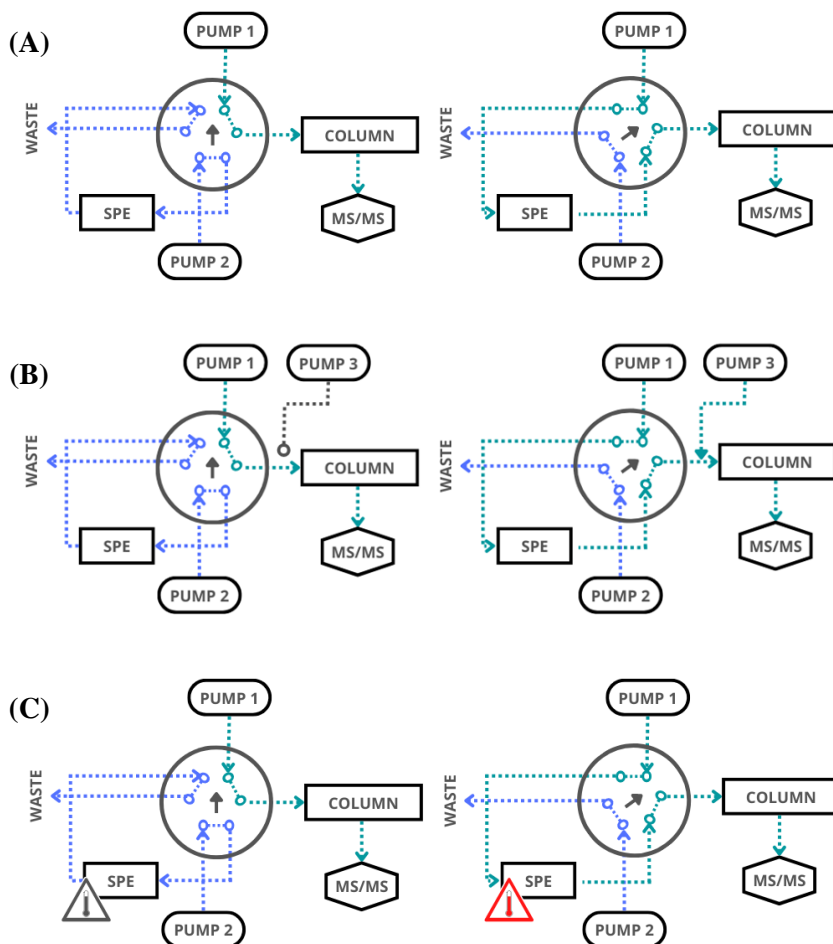
Figure S1 – On-line SPE configurations investigated in this study, illustrating loading (right) and injections (left) phases.

S4 Optimization of the chromatographic separation

Table S4 – Setups tested for the optimization of the chromatographic resolution (RChrom, see equation S1) of the isobaric groups (i) UR-A-GLU (A), iUR-A-GLU (B), RES-GLU (C) and (ii) QUE-3-GLU (D), QUE-7-GLU (E). For all the setups, the column was thermostatted at 40 °C and 0.1% formic acid was added in the eluents. SP = stationary phase, FR = flow rate (mL min^{-1}). Gradient elution: A = aqueous eluent and B = organic eluent; B percentage 2.0 % for 1 min followed by its increase at 2%/min for 13 min, 5%/min for 7.5 min, and final 100% for 3 min. In setups 9-11 acetonitrile was constantly added at 2% throughout the gradient.

Setup	Chromatographic conditions	R_{Chrom} A/B	R_{Chrom} A/C	R_{Chrom} B/C	R_{Chrom} D/E
1	SP: C18, FR: 0.4, A: H_2O, B: CH_3OH	0.0	0.4		0.0
2	SP: C18, FR: 0.6, A: H_2O, B: CH_3OH	0.0	0.5		0.0
3	SP: BP, FR: 0.4, A: H_2O, B: CH_3OH	0.0	2.4		0.0
4	SP: BP, FR: 0.4, A: H_2O, B: CH_3CN	2.6	0.7	1.9	0.5
5	SP: BP, FR: 0.5, A: H_2O, B: CH_3CN	2.7	0.8	2.0	0.6
6	SP: BP, FR: 0.6, A: H_2O, B: CH_3CN	2.6	0.6	1.9	0.0
7	SP: P-H, FR: 0.5, A: H_2O, B: CH_3CN	0.0	2.1		0.4
8	SP: P-H, FR: 0.6, A: H_2O, B: CH_3CN	0.0	2.2		0.5
9	SP: P-H, FR: 0.5, A: H_2O, B: CH_3OH	0.3	0.9	0.7	0.6
10	SP: P-H, FR: 0.6, A: H_2O, B: CH_3OH	0.7	1.4	0.9	0.8
11	SP: P-H, FR: 0.7, A: H_2O, B: CH_3OH	0.7	1.6	0.9	0.9

Equation S1 – Equation used for calculating chromatographic resolutions (R_{Chrom}) of the critical isobaric groups.

$$R_{Chrom} = \left(\frac{t_{R2} - t_{R1}}{\frac{(w_{b1} + w_{b2})}{2}} \right) \qquad (Eq.\ S1)$$
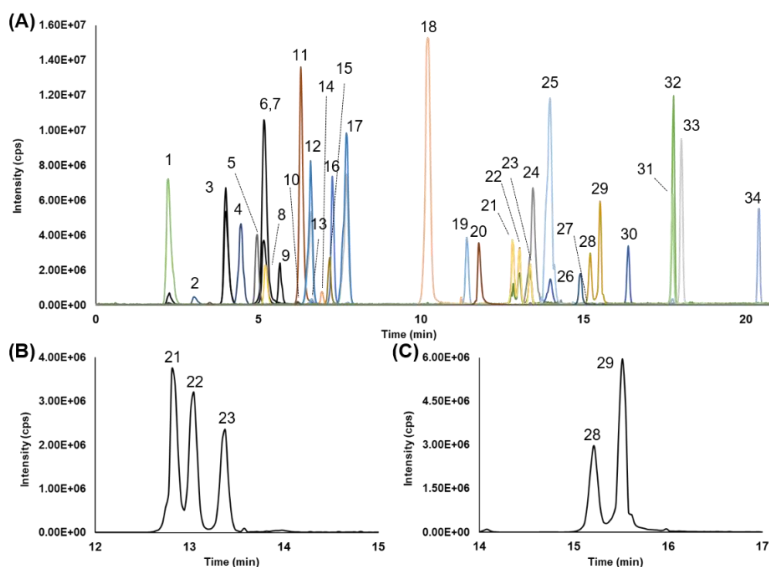

Figure S1 – (A) Complete MRM chromatogram of the 34 gut metabolites obtained with setup 11 and MRM chromatograms of (B) UR-A-GLU/iUR-A-GLU/TRES-GLU and (C) QUE-3-GLU/QUE-7-GLU. Injection volume = 50 µL. 1=DBA-SULF. 2=α-HHA, 3=2,5-DBA, 4=DCA-SULF, 5=CA-GLU, 6=2,3-DBA, 7=3-MC, 8=HA, 9=3-HPAA, 10=VA, 11=CA, 12=FA-SULF, 13=HVA, 14=FA-GLU, 15=CHA, 16=5-FQA, 17=iFA-SULF, 18=SAL, 19=FA, 20=UR-D, 21=UR-A-GLU, 22=iUR-A-GLU, 23=RES-GLU, 24=QUE-3-SULF, 25=TRES-SULF, 26=UR-C, 27=EA, 28=QUE-7-GLU, 29=QUE-3-GLU, 30=RES, 31=UR-B-GLU, 32=AA, 33=UR-A, 34=UR-B.

S5 On-line SPE optimization

Figure S3 – MRM chromatograms of four metabolites chosen as representative in terms of retention time window (see Figure S2) and polarity (log D at pH=3 in the range -0.77 – 2.50, see Table S2), obtained exposing the SPE cartridge to the elution gradient at room temperature. (A) CA-GLU, (B) 5-FQA, (C) 2,3-DBA and (D) UR-A.
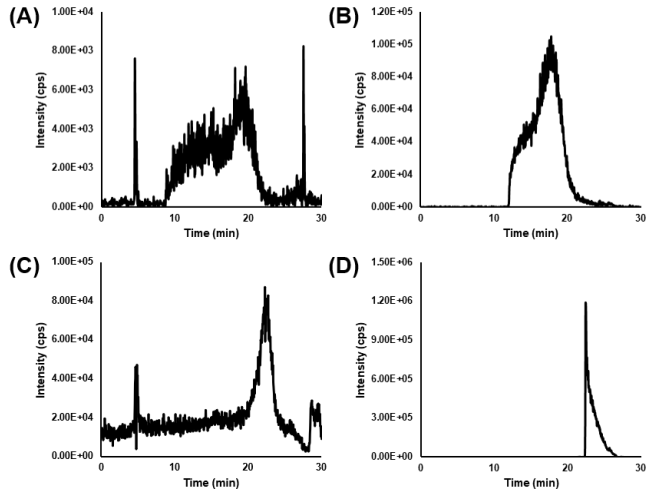

Figure S4 – MRM chromatograms of the four aforementioned metabolites, chosen as representative in terms of retention time window (see Figure S2) and polarity (log D at pH=3 in the range -0.77 – 2.50, see Table S2). The

chromatograms have been obtained under organic solvent assisted desorption and post-cartridge aqueous dilution (OD-PCD). (A) CA-GLU, (B) 5-FQA (C) 2,3-DBA, (D) UR-A.

Figure S5 – MRM chromatograms illustrating the critical resolutions UR-A-GLU/iUR-A-GLU (A-C-E) and QUE-3-GLU/QUE-7-GLU (B-D-F), obtained under OD-PCD (A-B), TD-system 1 (C-D), and TD system 2 (E-F).
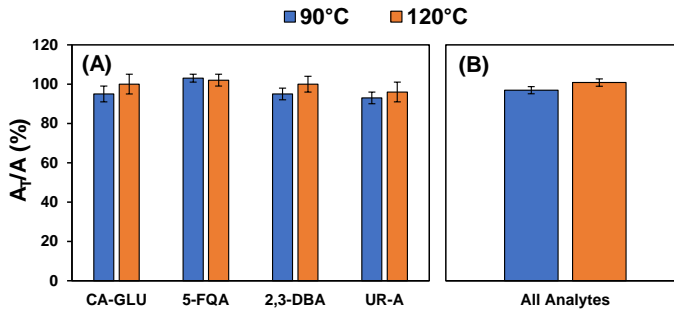
Figure S6 – Results of the thermal degradation tests expressed as percentage ratio between chromatographic areas after (A_T) and before (A) thermal exposure at 90

and 120 °C. Data referred to the four aforementioned metabolites (box A), chosen as representative in terms of retention time window (see Figure S2) and polarity (log D at pH=3 in the range -0.77 – 2.50, see Table S2) and to the mean value of all analytes (box B).
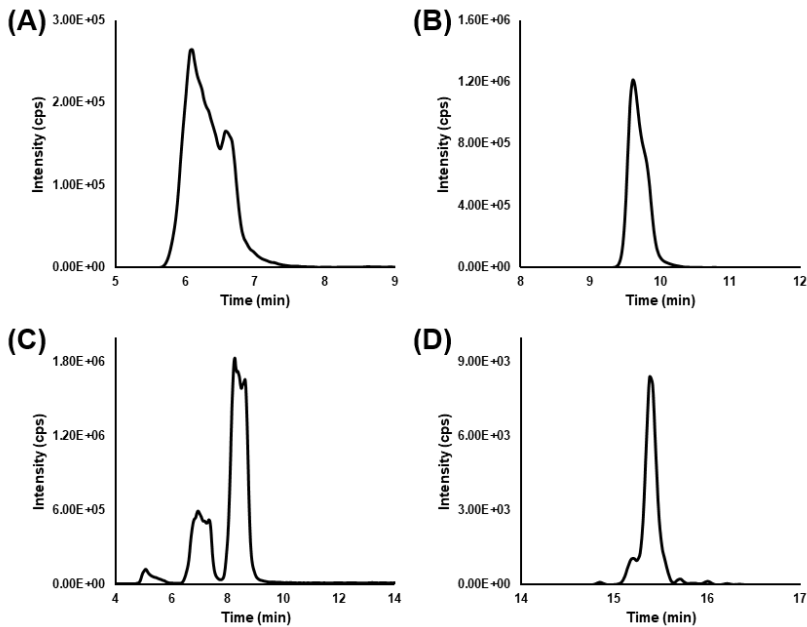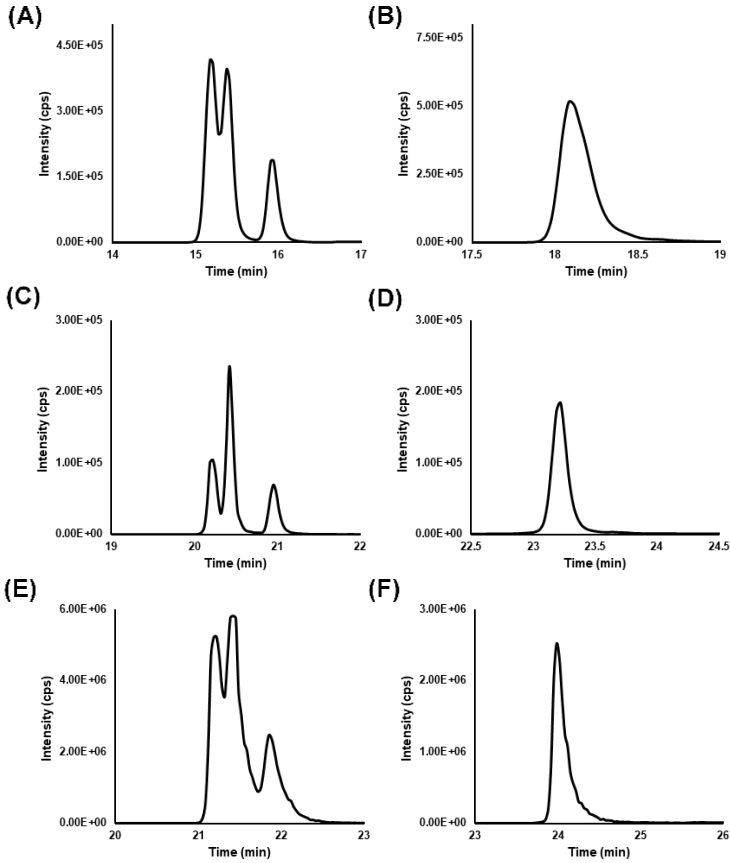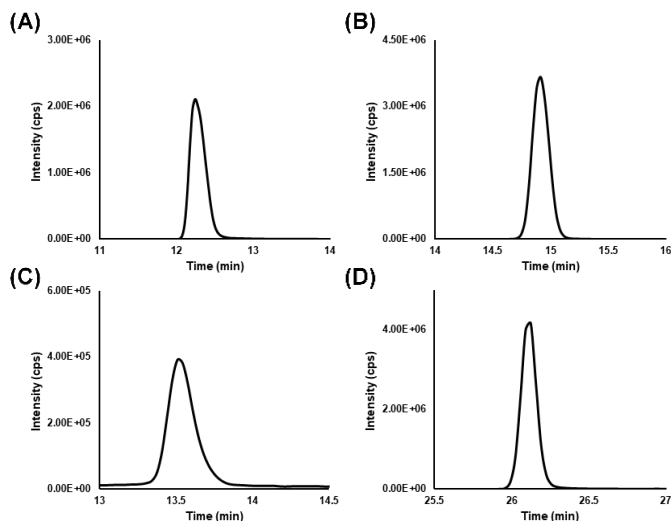

Figure S7 – MRM chromatograms of the four aforementioned metabolites, chosen as representative in terms of retention time window (see Figure S2) and polarity (log D at pH=3 in the range -0.77 – 2.50, see Table S2). The chromatograms have been obtained under thermal desorption with system 1. (A) CA-GLU, (B) 5-FQA (C) 2,3-DBA, (D) UR-A.

Figure S8 – MRM chromatograms of the four aforementioned metabolites, chosen as representative in terms of retention time window (see Figure S2) and polarity (log D at pH=3 in the range -0.77 – 2.50, see Table S2). The

chromatograms have been obtained under thermal desorption with system 2. (A) CA-GLU, (B) 5-FQA (C) 2,3-DBA, (D) UR-A.

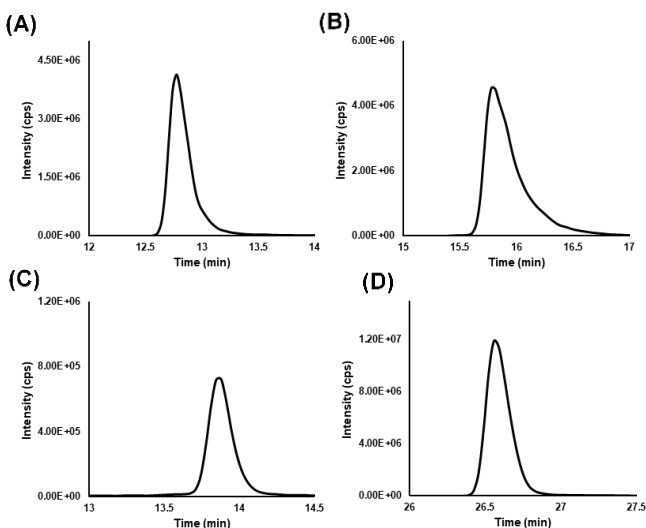**Table S5** – Instrumental figures of merit of the on-line SPE-TD-LC-MS/MS method for the 34 analytes. Instrumental detection (iLOD, µg/L) and quantitation (iLOQ, µg L^{-1}) limits, linearity range (µg L^{-1}), correlation coefficient (R^2), and mean (n=3) extraction efficiency (EE%) values, reporting intra-day reproducibility as RSD% in brackets.

Compound	iLOD	iLOQ	Linearity range	R^2	EE%
FA	0.04	0.12	0.12 - 500	0.9993	109 (3)
FA-GLU	0.13	0.43	0.43 - 500	0.9992	111 (3)
FA-SULF	0.03	0.11	0.11 - 500	0.9991	27 (2)
iFA-SULF	0.03	0.09	0.09 - 500	0.9970	58 (2)
VA	0.56	1.87	1.87 - 500	0.9996	75 (3)
HVA	0.10	0.34	0.34 - 500	0.9997	64 (2)
5-FQA	0.02	0.05	0.05 - 500	0.9969	111 (5)
SA	0.07	0.23	0.23 - 500	0.9976	93 (1)
CA	0.13	0.43	0.43 - 500	0.9977	113 (4)
CA-GLU	0.02	0.06	0.06 - 500	0.9994	110 (5)
DCA-SULF	0.01	0.03	0.03 - 500	0.9990	104 (6)
CHA	0.11	0.36	0.36 - 500	0.9997	113 (3)
EA	2.25	7.50	7.50 - 500	0.9993	115 (2)
2,3-DBA	1.68	5.60	5.60 - 500	0.9987	115 (4)
2,5-DBA	0.19	0.64	0.64 - 500	0.9973	85 (3)
3,4-DBA-SULF	0.02	0.05	0.05 - 500	0.9962	108 (2)
HA	0.10	0.33	0.33 - 500	0.9990	89 (6)
α-HHA	0.08	0.28	0.28 - 500	0.9990	65 (3)
3-HPAA	0.12	0.39	0.39 - 500	0.9996	61 (2)
AA	0.02	0.06	0.06 - 500	0.9978	115 (3)
3-MC	0.18	0.59	0.59 - 500	0.9984	70 (4)
QUE-GLU*	0.12	0.41	0.41 - 500	0.9988	110 (3)
QUE-3-SULF	0.11	0.38	0.38 - 500	0.9981	26 (2)
TRES	0.01	0.04	0.04 - 500	0.9995	107 (1)
TRES-GLU	0.09	0.30	0.30 - 500	0.9995	115 (4)
TRES-SULF	0.01	0.05	0.05 - 500	0.9998	110 (4)
UR-A	0.01	0.03	0.03 - 500	0.9989	106 (2)
UR-B	0.06	0.19	0.19 - 500	0.9969	102 (5)
UR-C	0.02	0.08	0.08 - 500	0.9987	105 (2)
UR-D	0.10	0.33	0.33 - 500	0.9984	104 (6)
UR-B-GLU	0.02	0.08	0.08 - 500	0.9978	96 (4)
UR-A-GLU	0.14	0.45	0.45 - 500	0.9989	107 (6)
iUR-A-GLU	0.15	0.50	0.50 - 500	0.9981	110 (5)
LUT**	0.02	0.07	0.07 - 500	0.9997	91 (2)
SIN**	0.09	0.30	0.30 - 500	0.9994	104 (6)

*QUE-3-GLU/QUE-7-GLU.

**Compounds used as internal standards for method validation.
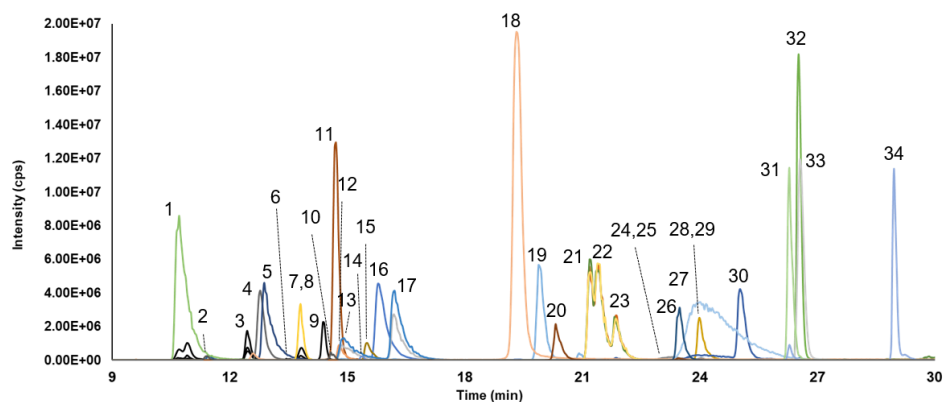
Figure S9 – Complete MRM chromatogram of the 34 target metabolites, illustrating the separation obtained with in the optimized on-line SPE-LC-MS/MS conditions. Injection volume = 500 µL. 1=DBA-SULF, 2=α-HHA, 3=2,5-DBA, 4=CA-GLU, 5=DCA-SULF, 6=3-MC, 7=HA, 8=2,3-DBA, 9=3-HPAA, 10=VA, 11=CA, 12=FA-GLU, 13=FA-SULF, 14=HVA, 15=CHA, 16=5-FQA, 17=iFA-SULF, 18=SA, 19=FA, 20=UR-D, 21=UR-A-GLU, 22=iUR-A-GLU, 23=RES-GLU, 24=QUE-3-SULF, 25=EA, 26=UR-C, 27=TRES-SULF, 28=QUE-7-GLU, 29=QUE-3-GLU, 30=TRES, 31=UR-B-GLU, 32=ABA, 33=UR-A, 34=UR-B.

S6 Method performances and sample analysis

Table S6 – Figures of merit of the on-line SPE-LC-MS/MS method in urine and serum samples. Matrix effect (ME%), apparent recovery (AR%) values and relative standard deviations (RSD%) at 25 µg L^{-1} (Level I) and 50 µg L^{-1} (Level II), mean recovery values (R%=AR%−ME), method detection (MDL, ng L^{-1}) and quantitation (MQL, ng L^{-1}) limits.

Compounds	*Urine*								*Serum*							
	ME%	Level I AR%	Level I RSD%	Level II AR%	Level II RSD%	R%	MDL	MQL	ME%	Level I AR%	Level I RSD%	Level II AR%	Level II RSD%	R%	MDL	MQL
FA	-14	92	7	96	4	108	43	128	13	117	2	120	3	106	34	101
FA-GLU	-23	87	1	88	2	111	149	491	12	50	3	70	3	48	217	717
FA-SULF	14	45	4	44	6	31	67	247	10	24	2	21	1	13	133	489
iFA-SULF	-11	50	5	49	3	61	61	182	13	36	3	26	1	18	97	290
VA	8	91	4	93	1	84	609	2033	-5	39	1	41	4	45	1400	4675
HVA	-6	61	3	60	1	67	165	562	21	33	1	50	3	21	241	819
5-FQA	-5	103	3	107	5	110	19	48	16	84	1	85	2	69	24	59
SA	-38	67	5	71	2	107	101	333	23	72	5	105	4	66	79	260
CA	-22	88	2	85	2	109	150	497	12	50	5	60	4	43	236	782
CA-GLU	2	101	1	98	0.4	98	20	60	33	65	3	71	4	35	29	88
DCA-SULF	12	114	5	113	1	102	9	26	24	56	1	63	3	36	17	50
CHA	8	113	1	115	2	106	96	316	14	40	2	53	3	33	237	774
EA	-14	102	8	105	1	118	2174	7246	16	35	1	30	3	17	6923	23077
2,3-DBA	-0.1	111	2	117	1	114	1474	4912	7	25	2	30	1	21	6109	20364
2,5-DBA	-9	77	1	75	3	85	250	842	12	20	5	46	4	21	576	1939
3,4-DBA-SULF	14	93	1	90	4	78	22	55	23	60	1	73	2	44	30	75
HA	-8	82	1	84	1	91	120	398	11	40	2	65	4	42	190	629
α-HHA	-2	67	5	66	4	69	120	421	28	41	4	60	4	23	158	554
3-HPAA	-8	54	1	57	2	64	216	703	12	33	5	40	5	25	329	1068
AA	-11	104	2	102	1	114	19	58	-13	83	5	90	5	100	23	69
3-MC	-1	65.4	0.5	62	1	65	283	926	15	29	1	35	4	17	563	1844
QUE-GLU*	-18	82.1	0.4	80	2	99	148	506	1	70	4	82	3	75	158	539
QUE-3-SULF	-7	29	1	30	2	37	373	1288	10	29	4	35	3	22	344	1188
TRES	-26	85	2	83	2	110	12	48	19	80	2	105	1	74	11	43
TRES-GLU	-5	110	1	108	1	114	83	275	19	80	1	105	2	74	97	324
TRES-SULF	-26	90	2	92	1	117	11	55	12	28	2	43	1	24	28	141

|															
UR-A	-9	101	1	100	2	110	10	30	-1	82	4	104	5	94	11	32
UR-B	-22	86	1	88	4	109	69	218	8	74	4	97	2	78	70	222
UR-C	-7	99	2	100	4	107	20	80	0.1	46	3	70	3	58	34	138
UR-D	16	88	1	83	2	70	117	386	4	64	1	77	4	67	142	468
UR-A-GLU	-0.3	105	1	103	1	104	135	433	5	62	1	80	3	66	197	634
UR-B-GLU	-12	82	8	84	2	95	24	96	4	106	5	102	6	100	19	77
iUR-A-GLU	1	111	3	109	4	109	136	455	4	119	1	115	2	113	128	427
LUT**	-24	62	1	61	3	86	33	114	-24	59	3	67	4	87	32	111
SIN**	-20	89	3	84	4	107	104	347	18	80	5	90	5	67	106	353

**QUE-3-GLU/QUE-7-GLU; **Internal standards
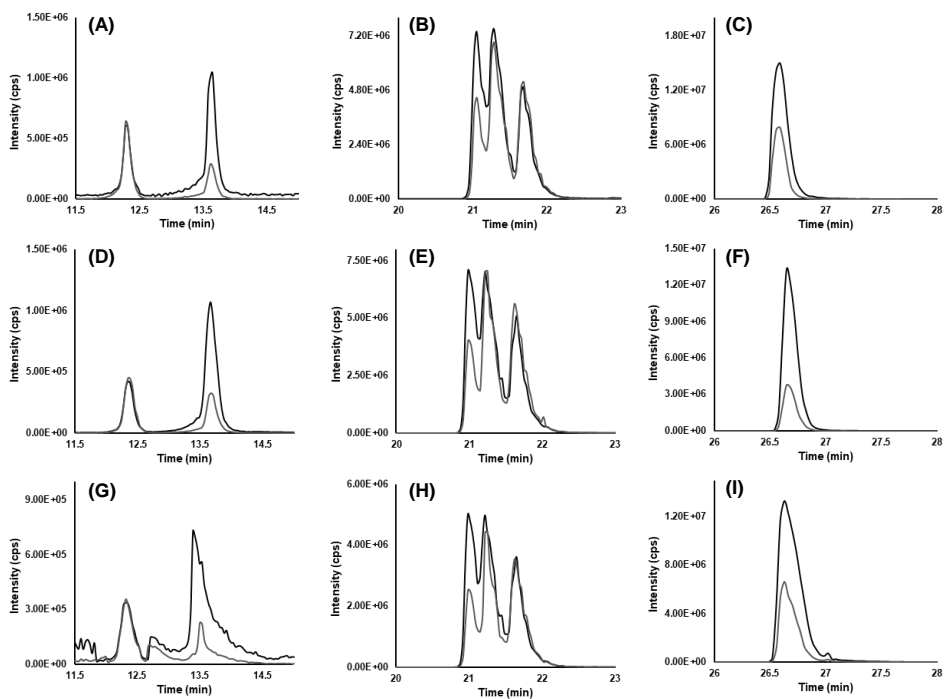
Figure S10 – Reconstructed multiple reaction monitoring chromatograms of 2,5DBA and 2,3DBA (A, D, and G), UR-A-GLU, iUR-A-GLU, and TRES-GLU (B, E, and H), and UR-A (C, F, and I) in spiked ultrapure water (A-C), urine (D-F), and serum (G-I).

9 General conclusions

Untargeted and targeted LC-MS/MS platforms are key tools capable to provide reliable analytical information about complex samples. Given the non-trivial and high dimensional (mainly in untargeted analysis) data structure, chemometrics and bioinformatic instruments can extract the latent information to address and identify the sample classification and origin, or the optimal experimental conditions for method development, or biological significance. Moreover, high throughput analytical technologies can be exploited to obtain more informative data on sample composition (e.g., LC×LC and on-line SPE).

In this PhD thesis, wet and dry lab LC-MS/MS platforms for the analysis of complex samples of environmental and nutritional interest have been presented. The key attributes for method development and optimization, as well as for sample characterization (i.e., targeted and untargeted LC-MS/MS features) have been pursued through modern and innovative LC-MS/MS technology combined with a proper conceptualization.

The presented untargeted platforms provided comprehensive and novel insights for chemotaxonomy and nutrimetabolomics by the annotation of relevant metabolites occurring in (poly)phenolic-rich fruits and diets. Data acquisition (i.e., DIA and DDA) and processing (e.g., non-parametric inference and computational metabolomics) strategies were combined to explore the metabolome space of interest and to increase its annotation coverage.

The analytical key features (e.g., selectivity and accuracy) of targeted platforms were exploited to retrieve quantitative information for (i) comparative analyses and (ii) analytical method optimization by multivariate approaches. Additionally, following the principles of pseudotargeted workflows, the relevant metabolite categories identified by the untargeted metabolome analysis of human biological fluids (i.e., serum and urine), were selected as model analytes for the development of a targeted high throughput analytical platform for their accurate quantification in urine samples.

Section 9

Concluding, the development and application of innovative wet and dry workflows have the power to improve the informative potential of the generated data, and consequently increase the knowledge on the chemical space of complex samples. Thereby, the scientific dissemination and the publicly available data provided in this PhD thesis work, could be considered as valuable contributions to the fulfilment of this analytical demand.