Bridging the gap between comprehensive extraction protocols in plant metabolomics studies and method validation

Reference:

Bijttebier Sebastiaan, Van der Auwera Anastasia, Foubert Kenn, Voorspoels Stefan, Pieters Luc, Apers Sandra.- Bridging the gap between comprehensive extraction protocols in plant metabolomics studies and method validation

Analytica chimica acta - ISSN 0003-2670 - 935(2016), p. 136-150

Full text (Publishers DOI): http://dx.doi.org/doi:10.1016/j.aca.2016.06.047

Bridging the gap between comprehensive extraction protocols in plant metabolomics studies and method validation

Authors: Sebastiaan Bijttebier[a,*], Anastasia Van der Auwera[a], Kenn Foubert[a], Stefan Voorspoels[b], Luc Pieters[a], Sandra Apers[a]

[a]University of Antwerp, Natural Products & Food Research and Analysis (NatuRA), Antwerp, Belgium. anastasia.vanderauwere@uantwerpen.be; kenn.foubert@uantwerpen.be; luc.pieters@uantwerpen.be; sandra.apers@uantwerpen.be

[b]Flemish Institute for Technological Research (VITO), Business Unit Separation and Conversion Technology (SCT), Mol, Belgium. stefan.voorspoels@vito.be

*Corresponding author:

Sebastiaan Bijttebier

Natural Products & Food Research and Analysis (NatuRA)

University of Antwerp

Universiteitsplein 1

2610 Antwerp

Belgium

Phone: +32 3 265 27 33

sebastiaan.bijttebier@uantwerpen.be

Abstract

It is vital to pay much attention to the design of extraction methods developed for plant metabolomics, as any non-extracted or converted metabolites will greatly affect the overall quality of the metabolomics study. Method validation is however often omitted in plant metabolome studies, as the well-established methodologies for classical targeted analyses such as recovery optimization cannot be strictly applied. The aim of the present study is to thoroughly evaluate state-of-the-art comprehensive extraction protocols for plant metabolomics with liquid chromatography-photodiode array-accurate mass mass spectrometry (LC-PDA-amMS) by bridging the gap with method validation.

Validation of an extraction protocol in untargeted plant metabolomics should ideally be accomplished by validating the protocol for all possible outcomes, *i.e*. for all secondary metabolites potentially present in the plant. In an effort to approach this ideal validation scenario, two plant matrices were selected based on their wide versatility of phytochemicals: meadow sweet (*Filipendula ulmaria*) for its polyphenols content, and spicy paprika powder (from the genus *Capsicum*) for its apolar phytochemicals content (carotenoids, phytosterols, capsaicinoids). These matrices were extracted with comprehensive extraction protocols adapted from literature and analysed with a generic LC-PDA-amMS characterization platform that was previously validated for broad range phytochemical analysis. The performance of the comprehensive sample preparation protocols was assessed based on extraction efficiency, repeatability and intermediate precision and on ionization suppression/enhancement evaluation.

The manuscript elaborates on the finding that none of the extraction methods allowed to exhaustively extract the metabolites. Furthermore, it is shown that depending on the extraction conditions enzymatic degradation mechanisms can occur. Investigation of the fractions obtained

with the different extraction methods revealed a low resolving power for phytochemicals for all methods. Nevertheless, an overall good repeatability was observed for all extraction methods, which is essential to allow direct comparison between samples. In summary, no single procedure outperforms the others and compromises will have to be made during method selection.

Keywords: comprehensive phytochemical extraction, plant metabolomics, liquid-chromatography-accurate mass mass spectrometry, validation

1. Introduction

Secondary plant metabolites have been found to have interesting applications as antioxidant, antitumor, cholesterol lowering, immunosuppressant, antiprotozoal, anthelminth and antiviral activities [1]. The assessment of the active constituents of a plant material is however often very complicated as plants contain a remarkably wide variety of metabolites; the total number of metabolites present in the plant kingdom is estimated at 200,000 or more [2]. Furthermore, synergistic interactions often play a vital role in explaining difficulties in isolating a single active ingredient and explaining the efficacy of apparently low doses of active constituents in a herbal product [3]. In addition, natural products are often pro-drugs which must undergo metabolic conversion before being active. Well known examples include the metabolic activation by gut bacteria of the phenolic compound salicin from willow bark to salicylic acid (anti-inflammatory agent) and the enzymatic hydrolysis of gluconasturtiin to phenethyl isothiocyanate (chemopreventive activity) upon crushing or masticating of cruciferous vegetables tissue [4, 5].

Until recently, most of these plant metabolites have been analysed with targeted methods for very specific purposes (e.g. quantitation of selected bioactive metabolites). With the advent of powerful modern analytical methods these metabolites can however be analysed more comprehensively [6]. The use of advanced analytical platforms such as ultra-high performance liquid chromatography-photodiode array-accurate mass mass spectrometry (UHPLC-PDA-amMS) and nuclear magnetic resonance (NMR) to characterize the metabolite composition (typically below 1500 Da) of a particular system or organism has been defined as metabolomics [7]. LC–MS is of particular importance for untargeted plant metabolomics studies due to this technique's sensitivity, selectivity, speed of structural elucidation and its capability of analysing a broad range of secondary metabolites such as alkaloids, benzoids, flavonoids, terpenes, isoprenes, glucosinolates and phenylpropanoids [7, 8].

Plant metabolomics differs from the traditional targeted phytochemical analysis in various fundamental aspects, such as being a data-driven approach with predictive power that aims to assess all measurable metabolites without any pre-conception or pre-selection [9]. Current extraction methods developed for plant metabolomics however do not allow collection of the entire plant metabolome, introducing inevitable compromises [10, 11]. Any non-extracted metabolites or metabolites converted by physical or chemical breakdown or enzymatic degradation will greatly affect the overall quality of the metabolomics study, which will be based on incomplete and skewed information [12-15]. Thus, it is vital to pay much attention to the extraction protocol design, and to understand and monitor its effects on the metabolic content and the profile obtained [12- 16]. Researchers however often ignore the importance of the extraction step, applying protocols based on their experience but omit to validate, or at least perform some evaluation of, their efficiency [14]. This is related to the fact that in such holistic analyses the well-established methodologies for recovery optimization or matrix effect evaluation for classical targeted analyses, for example, cannot be strictly applied [16].

At present, solvent extraction is most commonly used in metabolomics experiments [2]. Several efforts have been made in the past to develop a comprehensive extraction procedure for plant metabolites, e.g. through the use of combinations of different solvents for simultaneous extraction of many different plant compound classes [11, 16, 17-19]. The aim of the present study is to thoroughly evaluate these state-of-the-art comprehensive phytochemical extraction protocols for plant metabolomics with an UHPLC-PDA-amMS platform for broad range phytochemical analysis. Two plant matrices were selected for extraction method evaluation based on their wide versatility of phytochemicals: meadow sweet (*Filipendula ulmaria*) which is known for its polyphenols content, and spicy paprika powder (from the genus *Capsicum*) for its apolar phytochemicals content (carotenoids, phytosterols, capsaicinoids) [20, 21]. The performance of the comprehensive sample preparation protocols was assessed based on extraction efficiency, repeatability and

intermediate precision and on ionization suppression/enhancement evaluation, thereby bridging the gap with method validation.

2. Materials and methods

2.1. Chemicals

UHPLC-grade methanol, acetonitrile and ethyl acetate were purchased from Biosolve (Valkenswaard, The Netherlands). Ultrapure water with a resistivity of 18.2 MΩ·cm at 25 °C was generated with a Millipore system. Dichloromethane for gas chromatography, *n*-hexane for gas chromatography, acetone for gas chromatography and sodium hydrogen carbonate were purchased from Merck (Darmstadt, Germany). Formic acid, ammonium formate, ammonium acetate, (D-Ala^2)-leucine encephalin, butylated hydroxytoluene (BHT) and sand (quartz) were supplied by Sigma-Aldrich (Bornem, Belgium). Chloroform for HPLC was bought from Acros (Geel, Belgium). Commercially available mixtures to calibrate the mass spectrometer, *i.e.*, MSCAL5–1EA (caffeine, tetrapeptide ‘‘Met-Arg-Phe-Ala’’, Ultramark) for positive ion mode and MSCAL6–1EA (sodium dodecylsulfate, taurocholic acid sodium salt, Ultramark) for negative ion mode, were purchased from Supelco (Bellefonte, USA). Sulfuric acid and sodium hydroxide were supplied by Fisher Scientific (Loughborough, UK).

The following analytical standards were purchased from Phytolab (Vestenbergsgreuth, Germany): apigenin, luteolin, isorhamnetin, kaempferol, kaempferol-3-*O*-glucoside (astragalin), quercetin, quercetin-3-*O*-glucoside (isoquercitrin), quercetin-3-*O*-galactoside (hyperin), quercetin-3-*O*-rutinoside (rutin), quercetin-3-*O*-arabinoside (avicularin), quercetin-3-*O*-rhamnoside (quercitrin), galangin, phloretin, naringenin, (+)-catechin, (-)-epicatechin, (+)-dihydrokaempferol ((+)-aromadendrin), cyanidin-3-*O*-glucoside chloride (kuromanin chloride), cyanidin-3-*O*-rutinoside chloride (keracyanin chloride), procyanidin B2, ellagic acid and eriodictyol. Analytical standards of

salicylic acid, protocatechuic acid, gallic acid, ferulic acid, *p*-coumaric acid, caffeic acid, chlorogenic acid, β-carotene, *trans*-β-apo-8'-carotenal, γ-tocopherol, campesterol, β-sitosterol, miquelianin (quercetin 3-*O*-glucuronide), tannic acid and quinic acid were obtained from Sigma–Aldrich. Capsanthin, phytoene and phytofluene were purchased from Carotenature (Ostermundigen, Switzerland).

Filipendulae ulmariae herba (batch number 19969) was bought from Tilman SA (Baillonville, Belgium). Spicy paprika powder (99% paprika, chili pepper) from Verstegen Spices & Sauces N.V. was bought in a local store.

2.2. Preparation of standard solutions

Standard stock solutions for the phenolic analytes were prepared at a concentration of 1 mg mL-1 in UHPLC-grade methanol for each analyte separately and stored in the dark at 4 °C. Dilutions of these solutions were prepared in 60:40 (*v*:*v*) methanol:40 mM ammonium formate buffer (aqueous).

Standard stock solutions and working solutions for the non-phenolic analytes were prepared for each analyte separately at a concentration of approximately 200 µg mL-1. The stock solutions of phytosterols and lipid-soluble vitamins were prepared in methanol + 0.1% BHT. Stock solutions of carotenoids were prepared in dichloromethane + 0.1% BHT. Standard stock and working solutions were stored at −25 °C in the dark under an inert atmosphere (nitrogen). Dilutions of these solutions were prepared in dichloromethane + 0.1% BHT for analysis.

2.3. Sample preparation

Filipendula ulmaria was ground prior to extraction with an MF 10 basic Microfine grinder drive (IKA-Werke GmbH & Co. KG, Staufen, Germany) using a sieve mesh size of 0.5 mm. The spicy

paprika powder was extracted without grinding. All sample extractions were performed in triplicate unless indicated otherwise.

2.3.1. Reference extraction protocols

Polar reference extraction method: an extraction protocol previously developed by De Paepe et al. [22] was applied for the extraction of moderately polar phytochemicals. Briefly, 1 g of sample was extracted with 80:20 (*v:v*) water:methanol + 40 mM ammonium formate buffer (*v:v*) in a first step and 40 mM ammonium formate in methanol in a second step. Each extraction was performed by ultrasound-assisted solid–liquid extraction with 10 mL of the appropriate solvent by using a 2200 R-4 Ultrasonic sonicator (40 kHz, 100 W) (Branson Ultrasonic Corporation, Danbury, USA) for 60 min at room temperature. After 30 min of extraction, the solutions were vortex mixed (IKA MS2 Minishaker, IKAWerke GmbH & Co. KG). During sonication, the temperature was kept below 40 °C. Following the two consecutive extraction cycles, the supernatants were combined, centrifuged for 5 min at 3,000 rpm (approximately 1,450 g) using an Allegra™ Centrifuge (Beckman Coulter Inc., CA, USA), diluted 10 times with 60:40 (*v:v*) methanol:water + 40 mM ammonium formate and stored at 4 °C until analysis.

Apolar reference extraction method: a method previously developed by Bijttebier et al. [20] was used for the extraction of a wide array of apolar phytochemicals. Approximately 1 g of sample was mixed with approximately 1 g of sodium hydrogen carbonate and sand. Ultrapure water was added until the sample was hydrated (approximately 3 mL) and was let to rest in the dark under N_2 for 30 min to allow swelling of the matrix for better analyte extraction. Afterwards, the mixture was homogenized with sand and loaded into a 33 mL accelerated solvent extraction (ASE) cell (Thermo Fisher Scientific, Bremen, Germany). The mixture was extracted 3 times with 70:30 acetone:methanol + 0.1% BHT (*v:v*) with an ASE 200 (Thermo Fisher Scientific). The ASE settings

were as follows: oven temperature 40 °C, pressure 1050 psi, preheat 0 min, heat 5 min, static 2 min, flush percentage 65%, purge 100 s. The three extracts were combined in a separating funnel and 100 mL 10% NaCl (aqueous) and 15 mL hexane was added. The hexane phase was transferred to a recipient after vigorous shaking and the polar phase was extracted twice more with 15 mL hexane. The combined hexane fractions were evaporated to dryness, dissolved in a 10 mL solution of *trans*-β-apo-8'-carotenal (internal standard) in dichloromethane + 0.1% BHT and stored in the dark under nitrogen at -25 °C until analysis.

2.3.2. Comprehensive extraction protocols

Ethyl acetate extraction adapted from Halabalaki et al. [11]: 12.5 mL ethyl acetate was added to 1 g of sample. The mixture was shaken for 2 h followed by 1 h of ultrasound-assisted extraction. The mixture was subsequently centrifuged for 5 min at approximately 1,450 g. For analysis of polar metabolites, the extract was diluted 20 times with 60:40 (*v:v*) methanol:water + 40 mM ammonium formate and stored at 4 °C until analysis. For analysis of apolar metabolites, the extract was evaporated and dissolved in a 10 mL solution of *trans*-β-apo-8'-carotenal in dichloromethane + 0.1% BHT and stored in the dark under nitrogen at -25 °C until analysis.

Water:ethyl acetate extraction: the extraction procedure is similar to ethyl acetate extraction as described above, however, in addition 3 mL of water was added to the sample before extraction.

Chloroform:methanol:water extraction based on Theodoridis et al. [16] and Santos Pimenta et al. [17]: 2 mL of water, 4 mL of methanol and 4 mL of chloroform were added to 1 g of sample. The mixture was subsequently vortex mixed for 1 min followed by 5 min of ultrasound-assisted extraction, 15 min of shaking and 5 min of centrifugation at approximately 1,450 g. For analysis of polar metabolites, an aliquot of the upper phase was diluted 40 times with 60:40 (*v:v*)

methanol:water + 40 mM ammonium formate and stored at 4 °C before analysis. For analysis of apolar metabolites, the upper phase was removed and the lower apolar phase was collected. Additionally, 4 mL of chloroform was added to the remaining wet biomass and vigorously shaken for several seconds. The mixture was centrifuged for 5 min at approximately 1,450 g and the lower phase was combined with the previously collected apolar fraction. The combined chloroform extracts were evaporated, dissolved in a 10 mL solution of *trans*-β-apo-8'-carotenal in dichloromethane + 0.1% BHT and stored in the dark under nitrogen at -25 °C until analysis.

Continuous extraction adapted from Yuliana et al. [19]: a frit was inserted into an empty solid phase extraction column. Approximately 1 g of sample was mixed (not grinded) with 4 g of sea sand and loaded into the extraction column, followed by insertion of a frit on top of the sample mixture. The sample extraction column was subsequently mounted into a Reveleris iES flash instrument (Grace davison, Columbia, USA) and extracted with a continuous flow of solvents at 8 mL min-1. Fractions of 1 min (approximately 8 mL) were collected during the entire extraction cycle rendering 39 fractions (due to dead volume, no solvent is collected in fraction 1). The extraction solvents consisted of hexane (A), acetone (B), methanol (C) and water (D). The gradient for the continuous extraction from apolar to polar solvents (hereafter called ATP continuous extraction) was set as follows (min/A%/B%/C%): 0.0/100/0/0, 5.0/100/0/0, 15.0/0/100/0, 25.0/0/0/100, 35.0/0/0/0, 40.0/0/0/0. The gradient for the continuous extraction from polar to apolar solvents (hereafter called PTA continuous extraction) is the inverse of the above described gradient. For quantitative analysis of polar metabolites, the 25 most polar fractions were combined and diluted with methanol to 200 mL. For analysis of apolar metabolites, the 25 most apolar fractions were combined, evaporated, dissolved in a 10 mL solution of *trans*-β-apo-8'-carotenal in dichloromethane + 0.1% BHT and stored in the dark under nitrogen at -25 °C until analysis.

Polar to apolar accelerated solvent extraction (PTA ASE): approximately 1 g of sample was homogenized (mixed, not grinded) with sand and loaded into a 33 mL ASE cell (Thermo Fisher Scientific). The sample was extracted in 4 extraction cycles with 1: 50:50 (*v:v*) methanol:water + 40 mM ammonium formate, 2: methanol, 3: 50:50 (*v:v*) methanol:acetone, 4: acetone. The ASE settings were as follows: oven temperature 40 °C, pressure 1500 psi, preheat 0 min, heat 5 min, static 2 min, flush percentage 65%, purge 100 s. Each extraction cycle rendered 45 mL of extract. For analysis of polar metabolites, the extracts were diluted 5 times with 60:40 (*v:v*) methanol:water + 40 mM ammonium formate and stored at 4 °C until analysis. For analysis of apolar metabolites, all extracts were evaporated until only water remained or until dryness (depending on the solvent composition), and were subsequently extracted with or dissolved in a 10 mL solution of *trans*-β-apo-8'-carotenal in dichloromethane + 0.1% BHT and stored in the dark under nitrogen at -25 °C until analysis. The 4 different solvent fractions obtained for each sample were analysed separately. The areas obtained for the different fractions were summed for comparison with the other extraction protocols.

2.3.3. Evaluation of artefacts formation

Water extraction: hot water extraction was carried out by adding 10 mL of water at 90 °C to 1 g of sample, which was then placed in a hot water bath at 90 °C for 5 min. The sample was vortex mixed for 1 min and placed back into the hot water bath for 9 min. The sample was cooled in a water bath to room temperature and subsequently centrifuged for 5 min at approximately 1,450 g. After removal of the water extract, 10 mL of water at 90 °C was added to the sample followed by ultrasound-assisted extraction at 80 °C for 15 min. The sample was cooled in a water bath to room temperature, the water extracts were combined and centrifuged for 5 min at approximately 1,450 g. The extract was diluted 5 times with 60:40 (*v:v*) methanol:water + 40 mM ammonium formate

and stored at 4 °C until analysis. Water extraction at room temperature was carried out using the same extraction procedure at room temperature.

Methylation reaction adapted from Makkar [23]: 2 mL of methanol and 200 µL of sulfuric acid (≥ 95%) were added to separate vials containing 5 mg of gallic acid and ferulic acid, respectively. The vials were capped and stored for 20 h at 85 °C. After the solutions were cooled to room temperature, a solution of 0.1 g mL-1 NaOH in water was added until a pH between 5 to 6 was reached. These solutions were diluted with 60:40 (*v:v*) methanol:water, centrifuged for 5 min at approximately 1,450 g and analysed.

2.4. Instrumental analysis

The phytochemical composition of *Filipendula ulmaria* and chili peppers was previously characterized with two complementary LC-PDA-amMS protocols [20, 21]. During this study, analyses were performed using the same LC-PDA-amMS protocols. The extracted ion chromatograms (XICs) of the most abundant precursor ions of the phytochemicals were used for quantitative LC-amMS analysis. Detection of phytochemicals with the polar LC-PDA-amMS method was in general more sensitive in negative ion mode, while during analysis with the apolar LC-PDA-amMS method phytochemical detection was more sensitive in positive ion mode. Quantitation with the polar LC-PDA-amMS method was therefore performed in negative ion mode. Quantitation with the apolar LC-PDA-amMS method was performed in positive ion mode.

Analysis of moderately polar phytochemicals: 5 µL of extract was injected with a CTC PAL™ autosampler (CTC Analytics, Zwingen, Switzerland) on a Waters Acquity UPLC BEH SHIELD RP18 column (3.0 mm × 150 mm, 1.7 µm; Waters, Milford, USA) and thermostatically (40 °C) eluted with an Accela™ quaternary solvent manager and a 'Hot Pocket' column oven (Thermo Fisher Scientific).

The mobile phase solvents consisted of water + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B), and the gradient was set as follows (min/A%): 0.0/99, 9.91/74, 18.51/35, 18.76/0, 20.76/0, 20.88/99, 23.00/99. For detection, an amMS (Q Exactive™; Thermo Fisher Scientific) was used with heated electrospray ionization (HESI). Full scan data were acquired in negative ion mode over a mass/charge (*m*/*z*) range of 120-1800 and resolving power set at 70,000 at full width at half maximum (FWHM). Spray voltage was set at -2.5 kV, sheath gas and auxiliary gas at 47 and 15 (adimensional), respectively, and capillary temperature at 350 °C. The PDA detector was set to scan from 190 to 800 nm. Quality control during data acquisition with the generic LC-PDA-amMS method for polar phytochemicals was conducted by injecting a standard solution of a mixture of polyphenols before, between and after the sample extracts. The standard solution used for QC consisted of a mixture of phenolic and hydroxycinnamic acids, flavonoid aglycons and glycosides and tannins (compounds listed in section 2.1.). The compounds in this standard solution are native to *Filipendula ulmaria*: the standard mixture is therefore representative for the polar phytochemical composition of *Filipendula ulmaria*.

Analysis of apolar phytochemicals: 1.25 µL of extract was injected on a Waters Acquity UPLC HSS C18 SB column (2.1 mm × 100 mm, 1.8 µm; Waters) and thermostatically (35 °C) eluted. The mobile phase solvents consisted of 50:22.5:22.5:5 (*v*:*v*:*v*:*v*) water + 5 mM ammonium acetate:methanol:acetonitrile:ethyl acetate (A) and 50:50 (*v*:*v*) acetonitrile:ethyl acetate (B), and the gradient was set as follows (min/A%): 0.0/90, 0.1/90, 0.8/70, 20.0/9, 20.1/0, 20.4/0, 20.5/90, 23.0/90. Full scan data were acquired with atmospheric pressure chemical ionization (APCI) in positive ion mode over a *m*/*z*-range of 90-1400 and resolving power set at 70,000 at FWHM. The corona discharge current was set at 5 µA, the vaporizer and capillary temperatures were set at 450 °C. Lock mass correction with (D-Ala^2)-leucin enkephalin was applied. The PDA detector was set to scan from 190 to 800 nm. During data acquisition with the generic LC-PDA-amMS method for apolar phytochemicals, the internal standard was used to correct for instrumental fluctuations.

2.5. Method validation parameters

2.5.1. Process efficiency

During this study, the term process efficiency is used as described by Kruve et al. [24] (explained in **section 3.1.**). In this manuscript, the process efficiency is represented by the relative abundances of the phytochemicals per extraction method (Tables 1 and 2 in **section 3.1.** and **section 3.2.**). The relative abundances of the phytochemicals were calculated from the average areas of three replicates per extraction protocol, with the extraction protocol yielding the highest average area value set as 100 %.

2.5.2. Matrix effects on ionization efficiency

Changes in ionization efficiency caused by co-extracted matrix compounds were assessed by infusing a (D-Ala^2)-leucine encephalin solution. Sample extracts and solvent blanks were analysed with the LC-PDA-amMS methods while a 0.5 mg mL-1 solution of (D-Ala^2)-leucine enkephalin in methanol was infused post-column via a T-piece at a flow rate of 100 µL min-1. During the infusion experiments changes in ionization efficiency were investigated in both positive and negative mode.

2.5.3. Repeatability

The intra-day repeatability of the different extraction methods is represented by the standard deviations of three replicates per extraction method, prepared and analysed on the same day (Tables 1 and 2).

A more in-depth evaluation of the repeatability was carried out for the ATP continuous extraction protocol. Aliquots of *Filipendula ulmaria* and spicy paprika powder were extracted five times per day on three different days with the ATP continuous extraction protocol as described in **section 2.3.2.**. The repeatability and intermediate precision were determined simultaneously according to the Eurachem guidelines (Supplementary information 1) [25].

2.6. Strategy of experimental setup

2.6.1. Evaluation strategy

Without a deep understanding of the capabilities and limitations of the sample preparation method used in a given study, the accuracy of biological interpretation of collected data may be compromised [26]. Method validation is a key activity in chemical analysis, indispensable for obtaining reliable results [27]. As a rule, validation should become more important and voluminous with increasing complexity of the method [27]. Nonetheless, validation of sample preparation protocols in untargeted plant metabolomics studies is often omitted [14, 16]. This is because validation is difficult to achieve in untargeted analysis: how can you validate the unknown [28]? In our view, validation of an extraction protocol in untargeted plant metabolomics should ideally be accomplished by validating the protocol for all possible outcomes, *i.e*. for all secondary metabolites potentially present in the plant. In an effort to approach this ideal validation scenario, during this study we use generic LC-PDA-amMS methodology in combination with plants that contain a wide variety of phytochemicals to achieve an in-depth evaluation of comprehensive extraction protocols for plant metabolomics.

The use of standardized sample preparation protocols reduces technical variation and enhances reproducibility [11, 13, 29]. Moreover, combination of standardized sample preparation protocols with generic analytical methodology allows building compound libraries that enable effective

compound identification and efficient dereplication [10, 20]. In this view, a standardized generic characterization platform was previously developed for comprehensive phytochemical analysis in plants [20, 22]. This platform consists of two analytical methods complementary in terms of polarity: one sample preparation and UHPLC-PDA-amMS method for moderately polar compounds such as phenolic constituents [22] and another sample preparation and UHPLC-PDA-amMS method for apolar phytochemicals such as among others carotenoids and phytosterols [20]. A set of analytical standards representing the complex secondary metabolite composition present in nature (polyphenols (including flavonoid glycosides and aglycons, hydroxycinnamic and phenolic acids, procyanidins), carotenoids, fat soluble vitamins, phytosterols) was used to validate the applicability of the platform for broad range qualitative and quantitative analysis of phytochemicals in plant material [20, 22]. This platform makes use of an accurate mass MS detector for the tentative identification of unknown metabolites without the use of analytical standards: a hybrid quadrupole-orbital trap MS-analyser (Q Exactive™, Thermo Fisher Scientific) is used for selective ion fragmentation, a functionality that contributes significantly to compound identification by generating product ion spectra of selected precursor ions, thereby removing co-eluting interferences. PDA spectra, chromatographic behaviour, in-house and commercial compound databases and peer reviewed publications often provide additional confirmation of the proposed structures.

This methodology was previously applied to unravel the complex phytochemical composition of many plants, fruits and vegetables. It was shown before that two sample matrices, namely meadow sweet (*Filipenula ulmaria*) and red chili pepper (*Capsicum frutescens* L.), have a rich and complementary phytochemical composition [20, 21]. Meadow sweet, a medicinal plant used traditionally for, among others, its anti-inflammatory properties was shown to contain high amounts of moderately polar compounds such as phenolic acids, hydroxycinnamic acids, flavonoid aglycons and glycosides, and a rich collection of tannins. Chili pepper, a well-known spice, was found to contain high amounts of apolar phytochemicals (*i.e.* free and esterified carotenoids,

phytosterols, capsaicinoids and lipid-soluble vitamins). The phytochemicals of these two plants together cover a wide range of the chemical variety among secondary metabolites encountered in plants. These plants are therefore highly suited to be used as test matrices during the comparison of comprehensive extraction protocols for plant metabolomics studies.

During this study, meadow sweet and spicy paprika powder were extracted with several comprehensive extraction protocols and analysed with the two complementary UHPLC-PDA-amMS protocols. It is practically impossible to evaluate each identified phytochemical with a reference standard, in the first place as they are often not commercially available. Therefore, as an approximation, the areas of the individual constituents obtained with the validated sample preparation protocols of the generic characterization platform (**section 2.3.1.**)_were used as reference during evaluation. Keeping in mind that up to now no extraction method is able to exhaustively extract all metabolites, the results were critically evaluated when artefact formation or incomplete extraction was suspected.

2.6.2. Selection of comprehensive extraction methods

Halabalaki et al. [11] selected ethyl acetate as the most suitable solvent for investigating the variations in the composition of *Vitis* wood cultivars by untargeted metabolomics. The use of ethyl acetate enabled the extraction of a variety of secondary metabolites covering various polarities and chemical classes [11]. Notwithstanding its lipophilic character, ethyl acetate has also been shown to efficiently extract more polar compounds such as polyphenols during liquid-liquid extraction of wines [30, 31]. These findings indicate that ethyl acetate is an interesting solvent to test for broad range phytochemical extraction.

Several studies have however indicated that due to the presence of a wide array of metabolites in the plant kingdom, it is not possible to extract them with one single solvent [2, 13]. Efforts to ameliorate the simultaneous extraction of many different plant compound classes have been made

in the development of comprehensive sample preparation procedures by using combinations of different solvents [8]. Chloroform:methanol:water is a very useful mixture for the extraction of non-polar and polar compounds [2, 16, 17]. This solvent combination allows collection of apolar metabolites in the lower fraction and polar metabolites in the upper methanol:water fraction [16]. Kim et al. [13] reported that separate multivariate analyses of the two fractions provides much more information on the metabolites responsible for discriminating different samples of plants (different cultivars, treatments, etc.). Theodoridis et al. [16] investigated the optimal chloroform:methanol:water ratio for the extraction of metabolites from grapes. On the basis of the number of features (mass/retention time pairs) detected and other criteria such as phase separation, they selected 0.2:0.4:0.4 (*v:v:v*) H_2O:MeOH:$CHCl_3$ as the final extraction solvent composition [16].

Yuliana et al. [19] developed a continuous comprehensive extraction protocol in which a gradient of extraction solvents is used, ranging from apolar to polar (with solvents such as hexane, acetone, methanol, water). This protocol should in principle enable the extraction of the full range of metabolites. The continuous extraction protocol typically yields 20 to 30 fractions: qualitative analysis of each fraction (with partial least squares discriminant analysis and hierarchical cluster analysis) showed that the fractions are clustered in three to four groups based on their metabolite composition and that the extraction-fractionation protocol is repeatable [12, 18].

Accelerated solvent extraction (ASE) has been applied to a large number of botanicals and different anatomical plant parts including seeds, grains, herbs, rhizomes, processing by-products, etc. for the extraction of metabolites such as polyphenols and carotenoids [32-34]. Accelerated solvent extraction increases the efficiency of the extraction process by using conventional liquid solvents at elevated temperatures and pressures, which leads to lower solvent viscosity and better penetration into the matrix [35]. The application of this technique in a set-up similar to that of Yuliana et al. [19] could potentially ameliorate the extraction efficiency of phytochemicals.

The above described extraction protocols (extaction with ethyl acetate adapted from Halabalaki et al. [11], chloroform:methanol:water extraction, continuous extraction adapted from Yuliana et al. [19] and ASE) were selected for performance evaluation.

3. Results and discussion

3.1. Evaluation of comprehensive extraction protocols – polar phytochemicals

The results from the comparison of the extraction methods with the polar LC-PDA-amMS method are depicted in Table 1. As described by Kruve et al. [24], it is useful to make a distinction between recovery and process efficiency. Recovery refers to the losses and gains (caused by decomposition of other metabolites) of analyte during sample preparation, while process efficiency refers to loss or gain of analyte signal including the effects from sample preparation (recovery) and analyte ionization/detection [24]. Both process efficiency and the influence of matrix components on analyte ionization were investigated in this study. The relative abundances presented in Table 1 describe the process efficiency. Next to various phenolic derivatives and some triterpenes, other compounds such as amino acids, sugars and organic acids were also detected. As amino acids, sugars and organic acids are out of the scope of the manuscript, they are not included in Table 1.

The intensity drift during analysis calculated for the quality control samples was ≤ 15% for approximately 94% of the quality control injections, thereby confirming instrumental stability. A first view on the results in Table 1 shows that even for the limited polarity range that is investigated here, none of the tested extraction methods is able to exhaustively extract the metabolites. Nevertheless, an overall good repeatability was observed for all extraction methods, even when low relative abundances were encountered (an RSD ≤ 15% (n = 3) is obtained for approximately 96% of the compounds with a process efficiency ≥ 5%). The lowest process

efficiencies were obtained when the plant matrix was extracted with ethyl acetate. Only a small fraction of the compounds was extracted in comparison with the other extraction methods, except for several apolar triterpene derivatives which mark the apolar boundary of the polar LC-PDA-amMS method. These results are in contrast with the results of Halabalaki et al. [11] who reported that ethyl acetate was the most suitable solvent for extraction: during the current study dry *Filipendula ulmaria* herba was extracted, while during the study of Halabalaki et al. [11] *Vitis* wood samples were used for extraction solvent selection. Possibly the apolar character of ethyl acetate prohibits good permeation of the dry herbal *Filipendula* matrix. Michel et al. [34] compared different extraction solvents and also observed that hydrophilic compounds were more abundant in the methanol extracts of *Olea europaea* than in the ethyl acetate extract, although not to the same extent as was observed during the current study. Liquid-liquid extraction of wine with ethyl acetate has been used previously for the extraction of phenolic constituents [30, 31]. It was therefore tried during the current study to perform an extraction with ethyl acetate in combination with water. The results in Table 1 show that the addition of water leads to a large increase in the extractability of a select group of phenolic constituents: process efficiency increases for flavonoids in the order of diglycosides < monoglycosides < aglycons with regard to the other extraction methods. Most phenolic and hydroxycinnamic acids are also extracted in high amounts, while the majority of the tannins is not well extracted. As described by Naczk et al. [36], solubility of phenolic compounds is governed by several factors such as solvent used, as well as interaction of phenolics with other constituents, and there is no uniform procedure that is suitable for extracting all phenolics in plant materials. During this study it was observed that some of the phenolic constituents are even much better extracted with the water:ethyl acetate mixture than with the polar reference method. For instance, for kaempferol a relative abundance of only 38 % is obtained with the reference protocol, which is in contrast with the high recovery for kaempferol obtained with spiking experiments on dry apple matrix during method validation (92%) [22]. This discrepancy

in results might be caused by the difference in matrix composition or the fact that spiked compounds are not encapsulated in the matrix.

Analysis of the polar fraction of the chloroform:methanol:water extraction revealed that a much broader range of polar metabolites is extracted with this protocol than with ethyl acetate extraction (Table 1). Tannins and glycosylated flavonoids are well extracted. On the other hand, some of the phenolic acids, most of the flavonoid aglycons, the hydroxycinnamic acids and the triterpenes are only recovered in minor amounts. To check whether polar compounds did partition into the apolar chloroform fraction, the chloroform extracts were back-extracted with 2:1 (*v:v*) methanol:water. Analysis revealed that for approximately 60% of the compounds in Table 1 less than 5% of the total extracted amount was partitioned to the apolar phase. For approximately 21% of the compounds more than 15% of the total amount (ranging between 15% and 61%) was partitioned to the apolar phase. This group consisted mostly of flavonoid aglycons, hydroxycinnamic acids and triterpene derivatives, compounds of which only minor amounts are extracted with the chloroform:methanol:water extraction protocol. Partitioning of compounds into the apolar chloroform phase therefore only has minor influence on the relative abundances in Table 1.

The notion that the extraction method giving the highest signal intensity is optimum may not always be true, because high recovery of one metabolite may be caused by decomposition of other metabolites [26]. For instance, the physicochemical properties of methyl gallate and methyl caffeate are not that different from other phenolic compounds detected in *F. ulmaria*. Nonetheless, their relative abundance profile is very different from that of the other phenolic constituents. Relative to the polar reference extraction method, only low amounts of methyl gallate and methyl caffeate are obtained with the other extraction methods (Table 1). This suggests potential artefact formation during extraction with the reference method. It has been described

that in case of methanol as extraction solvent it is difficult to know whether a methoxy group is naturally occurring or is an artefact from the solvent [10, 37]. Thus, methyl gallate and methyl caffeate may potentially be formed by methylation of gallic acid and caffeic acid, both present in high abundance in *F. ulmaria* [21]. To investigate the source of these methylated acids, reference standards of gallic acid and caffeic acid were diluted in 80:20 (*v:v*) water:methanol + 40 mM ammonium formate and methanol + 40 mM ammonium formate. Aliquots of the dilutions were subsequently sonicated for 2 h. Analysis with LC-PDA-MS revealed no differences in composition before and after sonication, thereby excluding non-enzymatic methylation reactions. As an additional test, aliquots of *F. ulmaria* were extracted with the polar reference method, however, with the order of solvents in reverse (*i.e*. first with 100 % methanol + 40 mM ammonium formate to cause enzyme denaturation and secondly with 80:20 (*v:v*) water:methanol + 40 mM ammonium formate). Results showed that extraction with the same solvents in reverse order decreased the level of extracted methyl gallate and methyl caffeate to approximately 5 % and 2 %, respectively. This large decrease suggests that these compounds are formed by enzymes. To confirm these findings, ferulic acid, a compound which is not natively present in *F. ulmaria* but has a structure very similar to gallic acid and caffeic acid, was added before extraction to induce enzymatic methyl ferulate formation. Extraction of *F. ulmaria* according to the polar reference method in the presence of 5 mg ferulic acid did however only lead to the formation of minor amounts methyl ferulate. In contrast to the ratios of the areas of methyl caffeate/caffeic acid (4.0) and methyl ferulate/ferulic acid (4.7), the methyl ferulate/ferulic acid ratio remained very low (0.0004). Possibly methyl gallate and methyl caffeate are formed enzymatically out of the various galloyl and caffeoyl containing phenolic oligomers present in F. ulmaria [21], and not out of gallic acid and caffeic acid.

During the above described experiments, gallic acid and caffeic acid were identified with reference standards. However, no standards of methyl gallate, methyl caffeate and methyl ferulate were available so their identification during LC-PDA-amMS analysis remained tentative. To confirm their

presence, gallic acid and ferulic acid were methylated with a method adapted from Makkar [23]. LC-PDA-amMS analysis of the reaction products confirmed the identities of methyl gallate and methyl ferulate, thereby supporting the above hypothesis.

Inactivation of enzymes prior to extraction to avoid unpredictable artefact formation could be performed by using denaturing agents, heating or by using acids [6]. Care should on the other hand be taken to assure that the enzymatic inactivation procedure itself does not lead to artefact formation. For example, labile compounds such as carotenoids can degrade or isomerize under the influence of light, oxygen, enzymes, heat, oxidants, and acid or alkaline conditions. Carotenoids containing 5,6-epoxide moieties for instance easily undergo rearrangements to 5,8-epoxides, a reaction which is catalysed by acids [20]. Similarly, freeze-drying of tissue to remove water may lead to the irreversible adsorption of metabolites on cell walls and membranes [6].

The comprehensive extraction protocol developed by Yuliana et al. [19] consists of a continuous extraction with solvents ranging from apolar to polar. During the current study, extractions similar to that of Yuliana et al. [19] were carried out with solvents ranging from apolar to polar and vice versa, hereafter referred to as apolar to polar (ATP) and polar to apolar (PTA) continuous extraction, respectively (Table 1). Because of the large solvent volumes used during this continuous extraction process (approximately 300 mL), the constituents are heavily diluted. Nevertheless, the sensitivity of LC-PDA-amMS analysis allows to omit a concentration step. To investigate the process efficiency of moderately polar constituents with the ATP and PTA continuous extraction protocols, the 25 most polar fractions were combined, discarding the remaining apolar fractions. As will be discussed in **section 3.3.**, all compounds in Table 1 (with the exception of the apolar triterpenes) were predominantly recovered in the 25 most polar fractions of both the ATP and PTA continuous extraction. The relative abundances of the compounds extracted with the ATP and PTA continuous extraction were generally slightly elevated in comparison to those obtained in the polar fraction of the chloroform:methanol:water extraction. The relative abundances obtained with PTA continuous

extraction are similar to those obtained for ATP continuous extraction. However, in contrast with ATP continuous extraction, very low recoveries were found for salicylic primveroside derivatives monotropitin and spiraein (Table 1). Such low recoveries for primveroside derivatives were also observed for the polar reference extraction method, of which the first extraction solvent consists of 80:20 (v:*v*) water methanol + 40 mM ammonium formate (Table 1). It has been reported that monotropitin, also called gaultherin, present in the leaves of *Gaultheria procumbens* Linn. is hydrolysed in the presence of water by the enzyme gaultherase to produce methyl salicylate and a sugar, primverose [38]. To test whether the low recoveries for the primveroside derivatives encountered during the current study were due to interaction with enzymes, aliquots of *Filipendula ulmaria* were extracted with the polar reference method, however, with the order of solvents in reverse. This resulted in high process efficiencies for the primveroside derivatives, suggesting enzymatic degradation during the extraction methods starting with water or water:methanol mixtures. For confirmation, aliquots of *Filipendula ulmaria* were extracted with water at room temperature and with water heated to 90 °C (causing denaturation of enzymes). The extraction with water at 90 °C resulted in high process efficiencies for the salicylic primveroside derivatives, comparable to those obtained during ATP continuous extraction. The extraction with water at room temperature resulted in very low process efficiencies (0 – 2%). Flavonoid glycosides on the other hand were extracted well with both water at 90 °C and at room temperature, indicating that the difference in salicylic primveroside derivatives extraction is not temperature related but caused by enzymatic interactions. The reactivation of enzymes during extraction of fresh or dried plant material may lead to biased results: in the case of *Filipendula ulmaria*, part of the pharmacological effect can be explained by its salicylic acid component, which is released via oxidation from various aglycones (e.g. salicylaldehyde, methylsalicylate) developed from glycosides through hydrolysis in the digestive system [39]. Low extraction yields of salicylic primveroside derivatives might therefore lead to lower activity or false negative results.

Several method parameters were changed in an effort to optimize the process efficiency of the continuous extraction protocols, such as extraction with a lower flow rate (4 mL min-1 instead of 8 mL min-1), change in the steepness of the solvent gradient. None of the changes led to significant improvements. Continuous extraction with heated solvents could potentially lead to higher extraction efficiencies of phytochemicals. However, initial experiments with heated solvents revealed rapid and uncontrollable decreases in solvent temperature. Instrument tubings should therefore be isolated during future experiments to prevent these losses.

Extraction of *Filipendula ulmaria* with ASE was carried out sequentially with different solvents ranging from polar to apolar (PTA ASE). The relative abundances that were obtained for PTA ASE (Table 1) are in general slightly better than those of all other tested comprehensive extraction protocols. Moreover, no degradation of salicylic primveroside derivatives was observed. It was found previously during optimization of the apolar reference method that swelling of the matrix with water prior to ASE with an apolar extraction solvent caused a significant decrease in concentration of the C17-polyacetylenes falcarinol and falcarindiol [40]. Experimental data indicated that this decrease during sample swelling was most probably caused by enzymatic interactions [20]. On the contrary, the wetting step caused a significant increase in extracted carotenoids. So a trade-off may exist between unpredictable enzymatic breakdown and lower extraction yields. In a metabolomics experiment unpredictable changes to the metabolite composition are undesirable and therefore wetting of the sample should be avoided. To find out whether these effects caused by wetting of the sample prior to extraction also occurred with PTA ASE, *Filipendula ulmaria* was let to swell before PTA ASE: ultrapure water was added to the sample until it was hydrated (approximately 3 mL). Thereafter the sample was let to rest in the dark under N_2 for 30 min to allow swelling of the matrix. Analysis of the extracts revealed that for most compounds process efficiencies were comparable to those obtained with PTA ASE without sample wetting (data not shown). For hydroxycinnamic and phenolic acids and some flavonoid aglycons

however, higher extraction yields were obtained in comparison to PTA ASE without sample wetting. On the other hand, as with the polar reference method, very low recoveries were obtained for the salicylic primveroside derivatives, caused by enzymatic degradation during sample wetting. It should be tested in future experiments whether this wetting effect on compound recovery is matrix dependent. This was however not investigated during this study. Because of the higher pressures and temperatures applied during ASE, apolar to polar ASE without wetting of the sample prior to extraction could potentially result in process efficiencies that are better than those obtained for ATP and PTA continuous extraction, combined with lower solvent consumption.

In LC-MS analysis, ionization efficiency of the analytes may be strongly altered by co-eluting compounds which may cause a reduction or an increase of the analyte signal [27]. To find out whether (part of) the differences in relative abundances that were observed in Table 1 were caused by changes in ionization efficiency, a post-column infusion experiment was carried out with (D-Ala^2)-leucine encephalin (**section 2.5.2.**). Comparison of the (D-Ala^2)-leucine enkephalin intensity profile during analysis of the *Filipendula* extracts with that of a solvent blank revealed similar effects for the different sample preparation protocols: major ionization suppression and enhancement effects were detected in the beginning of the chromatogram (1 - 3 min). These effects, observed in both positive and negative ionization mode, are caused by early eluting polar bulk compounds such as sugars and organic acids. Only one of the secondary metabolites listed in Table 1, namely HHDP-hexoside, elutes in the 1 - 3 min region. Figure 1 shows a comparison of the (D-Ala^2)-leucine enkephalin intensity profile during analysis of a *Filipendula* extract from the polar reference method with that of a solvent blank (expressed as percent deviation). Due to instrumental fluctuations, variation in the (D-Ala^2)-leucine enkephalin could not be excluded. This prohibited investigation of minor changes in ionization efficiency caused by matrix effects. Nonetheless, as no other major ion suppression/enhancement effects were observed, the

differences in relative abundances in Table 1 are thus predominantly caused by differences in extraction efficiency.

The data in Table 1 show that the extraction efficiency changes depending on the solvents and techniques used: a non-compound-specific extraction protocol cannot be achieved. Oldiges et al. [41] therefore suggested that the basic requirement in a metabolomics experiment is the reproducibility of the protocol to allow direct comparison between samples, although the protocol potentially discriminates between chemical properties of the compounds. During the current study it was investigated whether an extraction method that allows only partial extraction of phytochemicals is quantitatively repeatable over several days. The repeatability and intermediate precision were calculated for the ATP continuous extraction protocol as described in **section 2.5.3.**. The intermediate precision data in Figure 2 show that the 15% boundary recommended by the U.S. Food and Drug Administration (FDA) regarding the analytical variability for targeted analysis, is only crossed for 5% of the compounds [42]. Although metabolomics is of a whole different fundamental analytical nature, the FDA guidance is used as a benchmark towards the repeatability evaluation of metabolomics approaches [15]. The repeatability data in Figure 2 demonstrate that even in the case of low extraction efficiencies for polar phytochemicals, extracts with highly similar composition are produced over different days with the ATP continuous extraction protocol. A thorough investigation of the stability of the phytochemicals in the (dried) extracts should be carried out in future experiments to determine a time frame in which the extracts may be used for metabolomics experiments.

3.2. Evaluation of comprehensive extraction protocols – apolar phytochemicals

Filipendula ulmaria only contains minor amounts of apolar phytochemicals [21]. As the genus *Capsicum* is a rich source of apolar phytochemicals [20], it was decided to use spicy paprika powder

consisting of a mixture of paprika (*Capsicum annuum)* and chili pepper (*Capsicum frutescens*) for in-depth evaluation of the performance of the comprehensive extraction protocols. The process efficiencies of selected comprehensive extraction procedures were compared with an apolar reference extraction method, previously developed for the generic extraction of apolar plant metabolites [20]. The results obtained from the comparison of the extraction methods with the apolar LC-PDA-amMS method are depicted in Table 2. A large variety of compounds such as carotenes, free and esterified xanthophylls, phytosterol aglycons and glycosides, capsaicinoids and a lipid soluble vitamin, was investigated. The relative abundances describe the process efficiencies.

The results in Table 2 show that also for apolar metabolites large differences in relative abundances are observed between extraction methods. Nonetheless, a good repeatability was obtained for the greater part of compounds (RSD ≤ 15% (n = 3) for approximately 80% of the compounds). High relative abundances were obtained for a broad spectrum of apolar metabolites during ethyl acetate and water:ethyl acetate extraction, which is in contrast with the low extraction efficiencies observed during polar phytochemical extraction (Table 1). Similar results were obtained during chloroform:methanol:water extraction. To check whether apolar compounds partitioned into the polar fraction during chloroform:methanol:water extraction, the polar fractions were back-extracted with a solution of *trans*-β-apo-8'-carotenal in dichloromethane + 0.1% BHT. Analysis revealed that the most polar compounds among the investigated apolar compounds, namely the capsaicinoids and glycosylated sterols, were partitioned only in small amounts to the polar phase (9% - 13% of the total extracted amount). The other apolar compounds were only partitioned for 2% or less into the polar phase.

The highest relative abundances for capsaicinoids and phytosterol glycosides were obtained for PTA ASE. In contrast, the lowest relative abundances for these compounds were encountered during extraction with the apolar reference method. This apolar reference method consists of three ASE cycles with 70:30 (*v:v*) acetone:methanol + 0.1% BHT followed by the addition of an

aqueous NaCl-solution and back-extraction with hexane, thereby favouring the extraction of apolar compounds.

To investigate the process efficiency of apolar phytochemicals with the ATP and PTA continuous extraction protocols, the 25 most apolar fractions were combined for analysis. It was noticed that in contrast with ATP continuous extraction, during PTA continuous extraction the greater part of capsaicinoids was recovered in the remaining polar fractions (78 – 88%). Other compounds were predominantly found in the 25 most apolar fractions during both ATP and PTA continuous extraction (≥ 96%). It was therefore decided to combine the results of the polar and apolar fractions of the PTA continuous extraction in the calculation of the relative abundances tabulated in Table 2. Still, higher relative abundances were found for capsaicinoids during ATP continuous extraction: these results may however be influenced by incomplete recovery of capsaicinoids during back-extraction of the remaining polar fractions with dichloromethane. Although the extraction solvents used during continuous extraction are also applied during PTA ASE (except hexane), lower relative abundances were obtained for capsaicinoids and phytosterol glycosides during continuous extraction. This is probably due to the differences in temperature and pressure applied during extraction. Lower relative abundances were observed for the compounds tentatively identified as antheraxanthin-laurate and antheraxanthin-myristate, and β-carotene (identified with analytical standard) during PTA ASE and PTA continuous extraction (with a large variability between replicates). The cause of these lower relative abundances should be investigated during future experiments. Nonetheless, the data in Table 2 show that none of the tested extraction methods is able to exhaustively extract the metabolites.

It is known that when using APCI high abundant matrix compounds such as lipids can cause major ionization suppression effects [43]. The ionization efficiency during analysis of the spicy paprika powder extracts was examined by post-column infusion of (D-Ala^2)-leucine encephalin as described in **section 2.5.2.**. No major ion suppression/enhancement effects were observed, indicating that

the differences in relative abundances in Table 2 are predominantly caused by differences in extraction efficiency. Nonetheless, as different ionization effects may be observed even for very similar matrices [44], ionization efficiency should ideally be evaluated for every new matrix.

It was investigated whether the extraction of apolar phytochemicals with the ATP continuous extraction protocol is quantitatively repeatable over several days. The experiments were conducted as described in **section 2.5.3.**. The results in Figure 3 show that the 15% boundary proposed by the FDA is only crossed for 3 compounds [42]. These data demonstrate that extracts with highly similar composition are produced over different days with the ATP continuous extraction protocol.

When combining the extraction efficiency data from Tables 1 and 2, it can be concluded that the extraction efficiency is dependent on the solvents and techniques used: none of the comprehensive extraction methods is non-compound specific. The most remarkable differences were observed for moderately polar phytochemicals. In contrast with other studies, ethyl acetate and water:ethyl acetate extraction resulted in very low extraction efficiencies for most polar compounds. With regard to the other tested extraction methods, the extraction efficiency of most compounds increases in the order of chloroform:methanol:water extraction < continuous extraction < PTA ASE. Nonetheless, in plants metabolomics studies other criteria besides extraction efficiency, such as ease of use, fractionation and dilution factor should also be considered. Chloroform:methanol:water is for instance easily scalable and only requires low amounts of solvent, while ASE and continuous extraction are less flexible in scalability and demand the availability of dedicated instrumentation. Furthermore, high amounts of organic solvents are consumed during continuous extraction, thereby resulting in heavily diluted fractions. During selection of the most appropriate extraction method, a compromise has to be made.

As described by Halabalaki et al. [11], either generic extraction procedures are used, or protocols are adapted to a given set of samples. A generic approach has the advantage of enabling comparisons between very different samples. A protocol adapted to a given sample type on the other hand has the advantage that detailed comparisons among a given set of samples can be obtained, as is required for the comparison of different cultivars [11]. The disadvantage of such a carefully adapted protocol is that the data generated are valid only for very specific samples and cannot be compared with other studies [11]. In the search for bioactive constituents of plants, it is important that the phytochemical composition of the extract is representative for the natural phytochemical composition of the plant material. Large differences in extraction efficiency between plant constituents could for example skew the results of pharmacological tests, leading towards wrong conclusions. In order to obtain an extract that represents the natural phytochemical composition of the plant material, extracts of different methods could be combined. The polar phytochemical composition of water:ethyl acetate and chloroform:methanol:water extraction is for example very complementary (Table 1). Combination of these extracts should result in an extract with a well-balanced phytochemical profile, *i.e.* an extract that is truly representative for the phytochemical composition native to the plant material under study.

3.3. Fractionation power of the extraction methods

In contrast with proteomics where enormous sample complexity demands depletion and fractionation steps, fractionation is not generally used in global LC–MS studies, while this could lead to deeper metabolome coverage and better distinction of the active metabolites [13, 26]. Several of the comprehensive extraction methods tested in the current study apply combinations of different solvents, rendering 2 or more fractions. As described in **sections 3.1.** and **3.2.**, fractionation during chloroform:methanol:water extraction resulted in two fractions with a clear separation of polar and apolar metabolites. During PTA ASE, 4 fractions are generated: most of the

moderately polar metabolites in Table 1 were extracted during the first extraction cycle (50:50 (*v:v*) methanol:water + 40 mM ammonium formate, fractionation data are shown in Supplementary information 2). Triterpenes were however predominantly recovered in the second fraction (100% methanol), because of their apolar character. Extraction of spicy paprika powder with PTA ASE resulted in the recovery of capsaicinoids in fraction 1, while other more apolar compounds such as carotenoids and phytosterols were predominantly found in fraction 2 (Supplementary information 3). Only small amounts of phytochemicals were encountered in fractions 3 and 4.

The continuous extraction protocol developed by Yuliana et al. [19] should in principle enable to acquire a more refined fraction composition as a linear gradient of 4 solvents is used, rendering 39 fractions. The distribution of the moderately polar phytochemicals from *Filipendula ulmaria* detected in the fractions obtained with the ATP continuous extraction protocol is tabulated in Supplementary information 2. The results are expressed as average relative abundance ± standard deviation of three extractions, of which each extraction was performed on a different day. The standard deviations in Supplementary information 2 demonstrate that fractions with highly similar composition are produced over different days with the ATP continuous extraction protocol. The highest abundance of most of the compounds during ATP continuous extraction is achieved in fractions 30 and 29 (methanol:water mixture). A second extraction maximum is found in fraction 20 (acetone:methanol mixture). In contrast with the other compounds, the apolar triterpenes are predominantly extracted in more apolar fractions 9-14 (hexane:acetone mixture). On the other hand, methyl gallate and methyl caffeate are detected in more polar fractions than the fractions containing the highest amounts of gallic acid and caffeic acid (Supplementary information 2). This is remarkable as methylation of phenolic and hydroxycinnamic acids renders them more apolar. This suggests that the small amounts of methyl gallate and methyl caffeate found are formed after extraction. The distribution of apolar phytochemicals from spicy paprika powder in the fractions obtained with the ATP continuous extraction protocol was examined (Supplementary information 3). The results were obtained as described for Supplementary information 2. The standard

deviations in Supplementary information 3 demonstrate that also for apolar compounds fractions with highly similar composition are produced with the ATP continuous extraction protocol over the different days. Most apolar compounds are predominantly collected in the first fraction. On the contrary, the more polar sterol glycosides are collected chiefly in fraction 9. Capsaicinoids are also collected, although to lesser extent, in and around fraction 9. When combining the fractionation data from Supplementary information 2 and 3, it can be concluded that ATP continuous extraction enables repeatable quantitative fractionation into 3 to 4 major groups with partly overlapping constituents.

If an objective of the sample preparation is to create fractions with different phytochemical composition, the question remains whether the comprehensive extraction-fractionation approach by Yuliana et al. [19] is the most appropriate option. Yuliana et al. [18, 19, 45] showed several times with proton NMR that this approach allows differentiation of the active constituents from the bulk of metabolites. However, the fractions obtained did not show a good resolution as many overlapping peaks were found [45]. Similar to the results during the current study, fractions could be divided in 3 to 4 major groups depending on the solvents used for extraction [18]. Yuliana et al. [45] suggest further optimization of the method by selecting solvent combinations that provide higher resolution of the metabolites over the fractions. This could however lower the extraction efficiency, which is already very low for certain compound groups with regard to other extraction protocols. Alternative approaches could be used that are more flexible and can generate a more orthogonal phytochemicals variation over the fractions. Classical preparative chromatography could for example be applied. One drawback of this approach is however that some compounds might not elute from the column. Moreover, if liquid chromatography is used, the products are usually collected at dilute conditions compared to the feed: concentrating the solute after collection not only requires the evaporation of significant quantities of solvent but may also lead to compound losses, such as the sublimation of salicylic acid during the evaporation of organic

solvents [46, 47]. Nonetheless, an evaporation/concentration step is equally required to render the sample suitable for pharmacological evaluation, metabolization studies or NMR analysis (unless deuterated solvents are used). Another option to create more phytochemicals variation over the fractions is to use a combination of sample preparation protocols with different selectivity in parallel, e.g. a comprehensive extraction protocol in combination with a more selective extraction protocol (e.g. in the case of *Filipendula ulmaria* water:ethyl acetate extraction).

4. Conclusions

Evaluation of the performance of extraction methods is often omitted in plant metabolomics. This can greatly affect the overall quality of the study. This study is the first to thoroughly evaluate and compare state-of-the-art comprehensive extraction protocols for untargeted plant metabolomics. This study fills an existing gap in the understanding of the role of sample preparation protocols in the outcome of plant metabolomics studies. Meadowsweet and spicy paprika powder were selected as matrices for extraction method comparison based on their versatile and complementary phytochemical composition (e.g. polyphenols, terpenes, capsaicinoids, phytosterols and carotenoids). Generic LC-PDA-amMS methodology developed for wide range phytochemical analysis allowed evaluating the comprehensive extraction methods on compound level: it was shown that none of the tested comprehensive extraction methods is able to exhaustively extract the phytochemicals. Furthermore, discrepancies in results with other studies, such as in the case of ethyl acetate extraction, show that phytochemical extraction is highly dependent on the plant matrix under study. It was also found that depending on the extraction conditions enzymatic activation can occur, thereby potentially leading to lower activity or false negative results. Nevertheless, an overall good repeatability was observed for all extraction methods, even when low relative abundances were encountered: this is essential to allow direct comparison between samples. Multiple criteria are important when selecting the most suitable

extraction method such as extraction efficiency, repeatability, fractionation power, ease of use, speed, dilution factor and scalability. Based on these criteria, the extraction method that best fits the goals and constraints of the metabolomics experiment as a whole should be selected. As shown during this study, no single procedure performs best in terms of all these criteria and a compromise will have to be made.

5. Acknowledgments

We thank the Research Foundation - Flanders (FWO) to support this research (projects 12M8315N and 12E5612N). Sebastiaan Bijttebier and Kenn Foubert are post-doctoral researchers of the FWO.

6. References

[1] P. Vaishnav, A.L. Demain, Unexpected applications of secondary metabolites, Biotechnology Advances 29 (2010) 223–229

[2] M. Ernst, D.B. Silva, R.R. Silva, R.Z.N. Vêncio, N.P. Lopes, Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing, Natural Product Reports 31 (2014) 784

[3] E.M. Williamson, Synergy and other interactions in phytomedicines, Phytomedicine 8 (2001) 401–409

[4] V. Butterweck, A. Nahrstedt, What is the best strategy for preclinical testing of botanicals? A critical perspective, Planta Medica 78 (2012) 747-754

[5] L.G. Wang, J.W. Chiao, Prostate cancer chemopreventive activity of phenethyl isothiocyanate through epigenetic regulation (Review), International Journal of Oncology 37 (2010) 533-539

[6] J.-L. Wolfender, S. Rudaz, Y.H. Choi, H.K. Kim, Plant Metabolomics: From Holistic Data to Relevant Biomarkers, Current Medicinal Chemistry 20 (2013) 1056-1090

[7] M. Castro-Puyana, M. Herrero, Metabolomics approaches based on mass spectrometry for food safety, quality and traceability, Trends in Analytical Chemistry 52 (2013) 74–87

[8] C.B. Hill, U. Roessner, Chapter 4: Advances in high-throughput untargeted LC–MS analysis for plant metabolomics, in: J. Han, C. Parker, C. Borchers (Eds.), Advanced LC-MS Applications in Metabolomics, Future Science Ltd, London, 2015, pp. 58-71

[9] F. Tugizimana, L. Piater, I. Dubery, Plant metabolomics: A new frontier in phytochemical analysis, South African Journal of Science 109(5/6) (2013) 1-11

[10] R. Verpoorte, Y.H. Choi, N.R. Mustafa, H.K. Kim, Metabolomics: back to basics. Phytochemistry Reviews 7 (2008) 525–537

[11] M. Halabalaki, S. Bertrand, A. Stefanou, K. Gindro, S. Kostidis, E. Mikros, L.A. Skaltsounis, J.-L. Wolfender, Sample Preparation Issues in NMR-based Plant Metabolomics: Optimisation for Vitis Wood Samples, Phytochemical Analysis 25 (2014) 350–356

[12] Y.H. Choi, R. Verpoorte, Metabolomics: what you see is what you extract, Phytochemical Analysis 25 (2014) 289–290

[13] H.K. Kim, R. Verpoorte, Sample Preparation for Plant Metabolomics, Phytochemical Analysis 21 (2010) 4–13

[14] M.Y. Mushtaq, Y.H. Choi, R. Verpoorte, E.G.Wilson, Extraction for Metabolomics: Access to The Metabolome, Phytochemical Analysis 25 (2014) 291–306

[15] R. t'Kindt, K. Morreel, D. Deforce, W. Boerjan, J. Van Bocxlaer, Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment, Journal of Chromatography B 877 (2009) 3572–3580

[16] G. Theodoridis, H. Gika, P. Franceschi, L. Caputi, P. Arapitsas, M. Scholz, D. Masuero, R. Wehrens, U. Vrhovsek, F. Mattivi, LC-MS based global metabolite profiling of grapes: solvent extraction protocol optimisation, Metabolomics 8 (2012) 175–185

[17] L.P. Santos Pimenta, H.K. Kim, R. Verpoorte, Y.H. Choi, Chapter 9. NMR-Based Metabolomics: A Probe to Utilize Biodiversity, in: U. Roessner, D.A. Dias (Eds.), Metabolomics Tools for Natural Product Discovery: Methods and Protocols, Methods in Molecular Biology, vol. 1055. Springer Science+Business Media Springer, New York Heidelberg Dordrecht London, 2013, pp 117-127

[18] N.D. Yuliana, 2011, Doctoral Thesis, Leiden University. Comprehensive extraction and NMR-based metabolomics: novel approaches to natural products lead finding in drug discovery. URL: https://openaccess.leidenuniv.nl/handle/1887/17704. Accessed on the 25th of January 2016

[19] N.D. Yuliana, A. Khatib, R. Verpoorte, Y.H. Choi, Comprehensive Extraction Method Integrated with NMR Metabolomics: A New Bioactivity Screening Method for Plants, Adenosine A1 Receptor Binding Compounds in *Orthosiphon stamineus* Benth, Analytical Chemistry 83 (17) (2011) 6902–6906

[20] S. Bijttebier, K. Zhani, E. D'Hondt, B. Noten, N. Hermans, S. Apers, S. Voorspoels, Generic Characterization of Apolar Metabolites in Red Chili Peppers (*Capsicum frutescens* L.) by Orbitrap Mass Spectrometry, Journal of Agricultural and Food Chemistry 62 (2014) 4812–4831

[21] S. Bijttebier, A. Van der Auwera, S. Voorspoels, B. Noten, N. Hermans, L. Pieters, S. Apers, A first step in the quest for the active constituents in *Filipendula ulmaria* (meadowsweet):

comprehensive phytochemical identification by liquid chromatography coupled to quadrupole-orbitrap mass spectrometry, Planta Medica, published online

[22] D. De Paepe, K. Servaes, B. Noten, L. Diels, M. De Loose, B. Van Droogenbroeck, S. Voorspoels, An improved mass spectrometric method for identification and quantification of phenolic compounds in apple fruits, Food Chemistry 136 (2013) 368–375

[23] H.P.S. Makkar, Quantification of Tannins in Tree and Shrub Foliage: A Laboratory Manual. Springer Science & Business Media, 2013, New York

[24] 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

[25] B. Magnusson, U. Örnemark (Eds.) Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (2nd ed. 2014). ISBN 978-91-87461-59-0. Available from www.eurachem.org Accessed on the 25th of January 2016

[26] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry, Analytical and Bioanalytical Chemistry 403 (2012) 1523–1548

[27] 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 I. Analytica Chimica Acta 870 (2015) 29–44

[28] G.A. Theodoridis, H.G. Gika, E.J. Want, I.D. Wilson, Liquid chromatography–mass spectrometry based global metabolite profiling: A review, Analytica Chimica Acta 711 (2012) 7– 16

[29] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based plant metabolomics: where do we stand, where do we go? Trends in Biotechnology 29(6) (2011) 267-275

[30] E. Porgalı, E. Büyüktuncel, Determination of phenolic composition and antioxidant capacity of native red wines by high performance liquid chromatography and spectrophotometric methods, Food Research International 45 (2012) 145–154

[31] Y. Ma, N. Tanaka, A. Vaniya, T. Kind, O. Fiehn, Ultrafast Polyphenol Metabolomics of Red Wines Using MicroLC-MS/MS. Journal of Agricultural and Food Chemistry 2016 64 (2) 505–512

[32] D.E. Breithaupt, Simultaneous HPLC determination of carotenoids used as food coloring additives: applicability of accelerated solvent extraction, Food Chemistry 86 (2004) 449–456

[33] P. Kraujalis, P.R. Venskutonis, A. Pukalskas, R. Kazernaviciut, Accelerated solvent extraction of lipids from *Amaranthus* spp. Seeds and characterization of their composition, LWT - Food Science and Technology 54 (2013) 528-534

[34] T. Michel, I. Khlif, P. Kanakis, A. Termentzi, N. Allouche, M. Halabalaki, A.-L. Skaltsounis, UHPLC-DAD-FLD and UHPLC-HRMS/MS based metabolic profiling and characterization of different *Olea europaea* organs of Koroneiki and Chetoui varieties, Phytochemistry Letters 11 (2015) 424–439

[35] F. Barros, L. Dykes, J.M. Awika, L.W. Rooney, Accelerated solvent extraction of phenolic compounds from sorghum brans, Journal of Cereal Science 58 (2013) 305-312

[36] M. Naczk, F. Shahidi, Extraction and analysis of phenolics in food, Journal of Chromatography A, 1054 (2004) 95–111

[37] M.M. Abou-Zaid, D.A. Lombardo, C. Nozzolillo, Methyl gallate is a natural constituent of maple (Genus Acer) leaves, Natural Product Research 23(15) (2009) 1373–1377

[38] M. Daniel, Medicinal plants - Chemistry and properties, CRC Press, Boca Raton, FL, 2006

[39] I. Papp, B. Simándi, B. Blazics, Á. Alberti, É. Héthelyi, É. Szöke, Á. Kéry, Monitoring Volatile and Non-Volatile Salicylates in *Filipendula ulmaria* by Different Chromatographic Techniques, Chromatographia Supplement 68 (2008) S125 – S129

[40] S. Bijttebier, E. D'Hondt, B. Noten, N. Hermans, S. Apers, V. Exarchou, S. Voorspoels, Automated analytical standard production with supercritical fluid chromatography for the quantification of bioactive C17-polyacetylenes: A case study on food processing waste, Food Chemistry 165 (2014b) 371–378

[41] M. Oldiges, S. Lütz, S. Pflug, K. Schroer, N. Stein, C. Wiendahl, Metabolomics: current state and evolving methodologies and tools, Applied Microbiology and Biotechnology 76 (2007) 495–511

[42] U.S. Food and Drug Administration, Guidance for Industry: Bioanalytical Method Validation, URL: http://www.fda.gov accessed at the 10th of February 2016

[43] S. Bijttebier, E. D'Hondt, B. Noten, N. Hermans, S. Apers, S. Voorspoels, Improving method reliability in carotenoid analysis through selective removal of glycerolipid interferences by lipase treatment, Journal of Agricultural and Food Chemistry 62 (14) (2014) 3114–3124

[44] H. Faccin, C. Viana, P.C. do Nascimento, D. Bohrer, L.M. de Carvalho, Study of ion suppression for phenolic compounds in medicinal plant extracts using liquid chromatography–electrospray tandem mass spectrometry, Journal of Chromatography A 1427 (2016) 111–124

[45] N.D. Yuliana, D. Rosa, S. Budijanto, R. Verpoorte, Y.H. Choi, Identification of adenosine A1 receptor ligands from *Morus alba* L. stem bark by NMR metabolomics approach, International Food Research Journal 21(3) (2014) 1067-1071

[46] A. Rajendran, Design of preparative-supercritical fluid chromatography, Journal of Chromatography A 1250 (2012) 227– 249

[47] M.C. Verberne, N. Brouwer, F. Delbianco, H.J.M. Linthorst, J.F. Bol, R. Verpoorte, Method for the Extraction of the Volatile Compound Salicylic Acid from Tobacco Leaf Material. Phytochemical Analysis 13 (2002) 45–50

Table 1. Relative abundances (%) of phytochemicals from *Filipendula ulmaria* obtained with the selected extraction protocols and analysed with the polar LC-PDA-amMS method. The results are expressed as average relative abundance ± standard deviation of three extractions.

	polar reference	ethyl acetate	water:ethyl acetate	chloroform:methanol: water polar fraction	ATP* continuous (25 most polar fractions)	PTA** continuous (25 most polar fractions)	PTA** ASE (summation of 4 fractions)
gallic acid	72 ± 2	< 1	100 ± 3	21.3 ± 1	22 ± 1	44 ± 5	26 ± 3
methyl gallate	100 ± 7	< 1	< 1	2.3 ± 0.2	4.6 ± 0.3	12 ± 1	5.7 ± 0.4
protocatechuic acid	100 ± 1	6.1 ± 0.4	84 ± 3	72 ± 3	79 ± 10	83 ± 3	91 ± 3
salicylic acid	88 ± 1	6.7 ± 0.4	79 ± 3	73 ± 2	80 ± 10	96 ± 4	100 ± 2
syringic acid	93 ± 3	< 1	3.7 ± 0.7	90 ± 2	100 ± 3	79 ± 4	80 ± 2
methyl caffeate	100 ± 5	< 1	< 1	< 1	< 1	2.1 ± 0.3	2.17 ± 0.09
caffeic acid	25.3 ± 0.7	< 1	100 ± 5	5.8 ± 0.3	7 ± 0.4	7.6 ± 0.7	6.3 ± 0.1
p-coumaric acid	25.1 ± 0.9	1.56 ± 0.07	100 ± 3	7.4 ± 0.2	10 ± 1	13.7 ± 0.7	12.9 ± 0.5
kuromanin	89 ± 1	1.7 ± 0.4	1.3 ± 0.1	81 ± 4	75 ± 4	84 ± 6	100 ± 2
catechin	83 ± 2	6.3 ± 0.6	66 ± 2	74 ± 3	75 ± 4	88 ± 5	100 ± 3
epicatechin	80.3 ± 0.8	6 ± 0.8	63 ± 3	69 ± 1	77 ± 2	84 ± 4	100 ± 4
epigallocatechin	100 ± 3	< 1	46.4 ± 0.6	87 ± 2	85 ± 2	88 ± 5	100 ± 3
quercetin	39 ± 3	1.3 ± 0.6	100 ± 2	3.8 ± 0.1	5 ± 0.3	15 ± 1	8.2 ± 0.1
naringenin	100 ± 8	8.5 ± 0.5	79 ± 3	16.3 ± 0.1	32 ± 2	57 ± 3	38 ± 1
aromadendrin	88 ± 5	5 ± 0.8	100 ± 4	23.8 ± 0.7	30.9 ± 0.6	50 ± 4	44 ± 2
eriodictyol	91 ± 9	3.7 ± 0.5	100 ± 5	19.4 ± 0.3	30 ± 3	50.6 ± 0.7	45 ± 2
luteolin	65 ± 5	5.1 ± 0.9	100 ± 7	18 ± 1	21 ± 2	34 ± 2	26.2 ± 0.7
methoxyflavonoid	57 ± 6	35 ± 3	100 ± 5	5.6 ± 0.6	25 ± 2	47 ± 3	41 ± 3
methoxyflavonoid	61 ± 2	11 ± 1	100 ± 5	15.5 ± 1	35 ± 3	50 ± 0.3	43 ± 3
kaempferol	38 ± 3	1.8 ± 0.2	100 ± 5	3 ± 0.2	4 ± 0.4	13 ± 1	6.2 ± 0.2
isorhamnetin	38 ± 2	3.7 ± 0.3	100 ± 8	9 ± 1	10 ± 1	22 ± 1	19 ± 1
ursolic acid	25 ± 2	59.1 ± 0.4	49 ± 1	5.3 ± 0.5	18 ± 3	80 ± 6	100 ± 7
tormentic acid	44.6 ± 0.5	72 ± 1	60 ± 2	8.3 ± 0.1	15 ± 2	58 ± 2	100 ± 3
pomolic acid	41 ± 2	72 ± 4	60 ± 7	13.8 ± 0.4	31 ± 2	61 ± 3	100 ± 2
2-pyrone-4,6-dicarboxylic acid	96 ± 2	2.6 ± 0.5	1.5 ± 0.02	78 ± 2	100 ± 2	74 ± 4	76 ± 2
ellagic acid	100 ± 3	8 ± 3	77 ± 4	46 ± 4	93 ± 3	73 ± 3	66 ± 8
HHDP-hexoside isomers	100 ± 4	1.5 ± 0.4	1.2 ± 0.1	34 ± 1	41 ± 1	47 ± 2	43 ± 3
HHDP-monogalloyl-hexoside isomers	100 ± 7	< 1	9.2 ± 0.4	15.1 ± 0.9	18.2 ± 0.7	28 ± 5	15 ± 2
galloyl-threonic acid isomers	100 ± 3	< 1	< 1	43 ± 3	64 ± 7	52 ± 1	50 ± 10
monogalloylhexoside isomers	100 ± 3	2.2 ± 0.4	2.4 ± 0.1	76 ± 3	84 ± 3	72 ± 3	83 ± 5
digalloylhexoside isomers	100 ± 3	2.2 ± 0.5	11.5 ± 0.5	65 ± 3	69 ± 2	64 ± 4	67 ± 5
trigalloylhexoside isomers	94 ± 2	5 ± 1	27.8 ± 0.8	94 ± 4	96 ± 4	85 ± 5	100 ± 7
caffeoyl-threonic acid	100 ± 2	2.4 ± 0.6	2.5 ± 0.1	76 ± 3	90 ± 7	78 ± 1	80 ± 10
galloyl-caffeoyl-threonic acid	100 ± 0.7	1.8 ± 0.4	18.9 ± 0.7	60 ± 4	74 ± 6	61 ± 1	55 ± 9
digalloyl-caffeoyl-threonic acid	100 ± 0.4	2.9 ± 0.6	23 ± 1	82 ± 5	89 ± 5	82 ± 1	80 ± 10
trigalloyl-caffeoyl-threonic acid	74 ± 3	< 1	9 ± 1	88 ± 6	89 ± 6	91 ± 3	100 ± 10
coumaroylthreonic acid isomers	87 ± 2	3.2 ± 0.7	8.9 ± 0.4	78 ± 3	86 ± 6	85 ± 4	100 ± 4

digalloyl-coumaroyl-threonic acid	93 ± 2	3.9 ± 0.9	40 ± 2	88 ± 5	91 ± 5	86 ± 2	100 ± 10
di-*O*-caffeylquinic acid isomers	100 ± 9	2.2 ± 0.3	55 ± 2	74 ± 6	83 ± 9	77 ± 8	90 ± 2
chlorogenic acid	94 ± 2	3.2 ± 0.6	4 ± 0.2	79 ± 3	96 ± 6	91 ± 2	100 ± 10
coumaroylquinic acid isomers	81 ± 2	3.5 ± 0.6	7.6 ± 0.4	80 ± 3	87 ± 4	84 ± 4	100 ± 2
pedunculagin isomer	64 ± 2	6 ± 1	3.8 ± 0.1	60 ± 5	96 ± 7	81 ± 2	71 ± 9
rugosin A	45 ± 2	2.9 ± 0.6	< 1	79 ± 7	100 ± 5	65 ± 1	70 ± 10
rugosin B	100 ± 5	2.6 ± 0.7	1 ± 0.1	64 ± 4	84 ± 3	88 ± 5	70 ± 10
rugosin D	60 ± 3	5.7 ± 1	7.1 ± 0.4	97 ± 2	100 ± 4	99 ± 5	87 ± 7
rugosin E	75 ± 3	4.9 ± 0.8	5.4 ± 0.3	83 ± 6	89 ± 2	100 ± 0.9	88 ± 10
tellimagrandin I	79 ± 3	6 ± 1	22 ± 1	76 ± 4	100 ± 5	65 ± 1	76 ± 7
tellimagrandin II	43 ± 0.9	5.1 ± 0.9	14 ± 2	88 ± 8	100 ± 2	49 ± 1	86 ± 7
casuarinin/casuarictin	58 ± 1	5 ± 1	18 ± 2	59 ± 4	100 ± 3	42 ± 2	66 ± 4
(epi)catechin coupled to $C_{15}H_{14}O_5$	92 ± 3	5.5 ± 0.8	58.7 ± 1	84 ± 2	77 ± 3	81 ± 4	100 ± 3
procyanidin dimer isomers	90 ± 2	5.8 ± 0.8	39 ± 1	84 ± 3	81 ± 3	85 ± 3	100 ± 3
procyanidin trimer isomers	99 ± 3	6.1 ± 0.9	28.1 ± 0.8	88 ± 3	81 ± 4	83 ± 5	100 ± 4
procyanidin tetramer isomers	100 ± 3	3.9 ± 0.6	12.9 ± 0.6	89 ± 5	81 ± 4	81 ± 4	92 ± 3
procyanidin pentamer isomers	89 ± 3	4.6 ± 0.8	7.2 ± 0.5	81 ± 2	73 ± 3	86 ± 4	100 ± 2
procyanidin hexamer isomers	93 ± 3	3 ± 0.8	10.7 ± 0.2	83 ± 1	75 ± 2	85 ± 4	100 ± 3
procyanidin heptamer isomers	100 ± 6	< 1	< 1	80 ± 3	81 ± 3	92 ± 4	97 ± 5
procyanidin dimer gallate	100 ± 2	4.6 ± 0.9	58 ± 1	89 ± 5	80 ± 4	81 ± 3	95 ± 4
isosalicin	76 ± 1	2.9 ± 0.3	5.2 ± 0.3	80 ± 4	88 ± 6	83 ± 4	100 ± 3
spiraein	2.5 ± 0.1	5.2 ± 0.6	< 1	43 ± 3	85 ± 3	21 ± 6	100 ± 2
monotropitin	2.3 ± 0.09	5.2 ± 0.7	< 1	26 ± 2	57 ± 2	7 ± 3	100 ± 3
salicylic acid hexoside	94 ± 2	4.7 ± 0.6	< 1	92 ± 4	97 ± 3	84 ± 4	100 ± 2
rutin	69.5 ± 0.5	3.1 ± 0.5	2.5 ± 0.3	75 ± 2	70 ± 4	84 ± 4	100 ± 1
quercitrin	72 ± 2	5.2 ± 0.5	63 ± 2	65 ± 3	61 ± 3	68 ± 4	100 ± 3
hyperoside + isoquercitrin	64 ± 3	3.8 ± 0.5	30 ± 1	61 ± 1	59 ± 2	87 ± 4	100 ± 2
astragalin	79.9 ± 0.4	5.8 ± 0.7	50 ± 2	73 ± 2	77 ± 3	100 ± 5	99 ± 2
miquelianin	100 ± 2	3.1 ± 0.6	3.3 ± 0.1	75 ± 3	90 ± 8	83 ± 5	86 ± 7
quercetin-*O*-galloyldihexoside	83 ± 1	3 ± 0.6	14.8 ± 0.5	75 ± 2	64 ± 1	91 ± 6	100 ± 2
isorhamnetin-*O*-hexoside isomer	100 ± 3	5 ± 1	46 ± 3	90 ± 10	95 ± 5	98 ± 7	57 ± 3
methoxyflavonoid-*O*-hexoside-deoxyhexoside	72.8 ± 0.7	11 ± 1	4.3 ± 0.2	60.8 ± 0.4	63 ± 3	70 ± 3	100 ± 2
quercetin-*O*-galloylhexoside isomers	100 ± 1	4.5 ± 0.9	64 ± 1	87 ± 3	77 ± 3	85 ± 5	99 ± 3
kaempferol-*O*-hexoside-deoxyhexoside	62 ± 3	3.9 ± 0.6	4.2 ± 0.6	68.4 ± 0.9	67 ± 3	73 ± 4	100 ± 2
quercetin-*O*-pentoside	82.9 ± 0.2	5.1 ± 0.7	61 ± 2	70 ± 1	71 ± 3	82 ± 5	100 ± 6
isorhamnetin-*O*-hexoside isomer	78 ± 2	3.1 ± 0.1	52 ± 4	60 ± 2	55 ± 1	71 ± 4	100 ± 2
flavonoid-*O*-galloylhexoside	97 ± 2	4.2 ± 0.8	80 ± 3	83 ± 3	73 ± 2	84 ± 2	100 ± 3
quercetin-*O*-digalloylhexoside isomers	92 ± 1	3.4 ± 0.9	80 ± 1	86 ± 3	70 ± 3	80 ± 3	100 ± 4
kaempferol-*O*-hexoside	50 ± 2	5 ± 0.7	39 ± 2	55 ± 1	65 ± 2	73 ± 2	100 ± 3
methoxyflavonoid-*O*-hexoside-deoxyhexoside	81 ± 5	10 ± 1	14.5 ± 0.1	54 ± 5	58 ± 2	90 ± 6	100 ± 2
dimethoxyflavonoid-*O*-hexoside-deoxyhexoside	93 ± 3	9.3 ± 0.8	14.6 ± 0.1	60 ± 5	64 ± 4	96 ± 10	100 ± 4
quercetin-*O*-galloyldeoxyhexoside	96 ± 2	4.4 ± 0.6	91 ± 1	83 ± 4	65 ± 4	83 ± 3	100 ± 3
avicularin/avicularoside	84 ± 2	5.3 ± 0.8	63 ± 2	67 ± 2	68 ± 3	83 ± 5	100 ± 5
aesculetin	68 ± 4	2.9 ± 0.5	74 ± 3	28 ± 2	43 ± 4	57 ± 2	100 ± 10

*ATP: apolar to polar

**PTA: polar to apolar

Table 2. Relative abundances (%) of phytochemicals from spicy paprika powder obtained with the selected extraction protocols and analysed with the apolar LC-PDA-amMS method. The results are expressed as average relative abundance ± standard deviation of three extractions.

	apolar reference	ethylacetate	water:ethyl acetate	chloroform:methanol: water apolar fraction	ATP* continuous extraction (25 most apolar fractions)	PTA** continuous extraction (summation of polar and apolar fraction)	PTA** ASE (summation of 4 fractions)
nordihydrocapsaicin	5.3 ± 0.7	55 ± 2	75 ± 5	64 ± 6	74 ± 2	44 ± 4	100 ± 7
capsaicin	5.2 ± 0.6	54 ± 4	77 ± 4	60 ± 7	67 ± 1	41 ± 4	100 ± 4
dihydrocapsaicin	9.6 ± 0.7	58 ± 2	77 ± 5	70 ± 4	75 ± 3	35 ± 5	100 ± 8
homodihydrocapsaicin	15 ± 2	61 ± 4	75 ± 5	64 ± 6	78 ± 3	27 ± 9	100 ± 8
cycloviolaxanthin	90 ± 4	89 ± 1	92 ± 4	84 ± 9	89 ± 4	45 ± 5	100 ± 6
campesteryl-glycoside	22.5 ± 0.8	45 ± 3	60 ± 3	47 ± 4	52 ± 2	40 ± 10	100 ± 5
cis-capsanthin	47.8 ± 1	61 ± 3	68 ± 5	50 ± 4	43 ± 2	47 ± 8	100 ± 6
capsanthin	92 ± 3	84 ± 4	90 ± 10	90 ± 7	97 ± 6	70 ± 4	100 ± 8
β-sitosteryl-glycoside	28.7 ± 0.8	46 ± 2	57 ± 3	49 ± 4	48 ± 2	40 ± 10	100 ± 9
antheraxanthin	98 ± 8	96 ± 3	99 ± 9	90 ± 10	100 ± 2	63 ± 6	70 ± 20
γ-tocopherol	89 ± 6	76 ± 2	78 ± 8	80 ± 10	79 ± 4	90 ± 10	100 ± 5
campesterol	74 ± 3	61 ± 4	62 ± 5	73 ± 5	71 ± 1	70 ± 10	100 ± 9
capsorubin-myristate	97 ± 5	89 ± 7	91 ± 9	100 ± 10	95 ± 5	70 ± 10	60 ± 20
capsanthin-laurate isomer	100 ± 8	85 ± 7	81 ± 7	100 ± 10	95 ± 7	90 ± 10	60 ± 10
β-sitosterol	74 ± 7	61 ± 2	65 ± 1	73 ± 7	71 ± 4	63 ± 10	100 ± 10
capsanthin-laurate isomer	100 ± 6	81 ± 3	85 ± 6	90 ± 10	90 ± 10	85 ± 6	79 ± 9
antheraxanthin-laurate	89 ± 3	99 ± 7	100 ± 10	80 ± 10	94 ± 3	50 ± 40	20 ± 20
capsanthin-myristate isomer	97 ± 9	95 ± 6	100 ± 8	87.8 ± 0.2	100 ± 20	90 ± 20	70 ± 20
β-carotene	98 ± 4	100 ± 5	100 ± 6	84 ± 9	94 ± 2	70 ± 30	40 ± 10
phytofluene	96 ± 3	86 ± 5	85 ± 5	91 ± 6	91 ± 5	90 ± 10	100 ± 6
capsanthin-myristate isomer	100 ± 20	93.8 ± 0.9	100 ± 9	86 ± 8	92 ± 10	80 ± 20	70 ± 20
antheraxanthin-myristate	95 ± 4	100 ± 5	100 ± 20	90 ± 20	100 ± 8	40 ± 40	20 ± 20
phytoene	79 ± 2	66 ± 3	64 ± 5	71 ± 9	76 ± 3	80 ± 10	100 ± 5
capsanthin-di-laurate	88 ± 7	83 ± 1	81 ± 4	90 ± 10	100 ± 6	90 ± 20	70 ± 7
capsanthin-laurate-myristate	100 ± 10	87 ± 5	90 ± 10	96 ± 8	100 ± 10	90 ± 20	70 ± 8

*ATP: apolar to polar

**PTA: polar to apolar

Figure captions

Figure 1. A comparison of the (D-Ala^2)-leucine enkephalin intensity profile during analysis of a *Filipendula* extract from the polar reference method with that of a solvent blank (expressed as percent deviation). Major changes in ionization efficiency caused by matrix compounds are restricted the first 3 min of the analysis.

Figure 2. Repeatability and intermediate precision of phytochemicals from *Filipendula ulmaria* obtained with the ATP continuous extraction, calculated according to the Eurachem guidelines [25].

Figure 3. Repeatability and intermediate precision of phytochemicals from spicy paprika powder obtained with the ATP continuous extraction, calculated according to the Eurachem guidelines [25].

Supplementary information captions

Supplementary information 1. Formulas used for the calculation of the repeatability and intermediate precision according to the Eurachem guidelines [25].

Supplementary information 2. The distribution of compounds from Filipendula ulmaria detected in the fractions obtained with: left: ATP continuous extraction protocol. The results are expressed as average relative abundance ± standard deviation of three extractions, of which each extraction was performed on a different day. Right: PTA ASE protocol. The results are expressed as average relative abundance ± standard deviation of two extractions.

Supplementary information 3. The distribution of compounds from spicy paprika powder detected in the fractions obtained with: left: ATP continuous extraction protocol. The results are expressed

as average relative abundance ± standard deviation of three extractions, of which each extraction was performed on a different day. right: PTA ASE protocol. The results are expressed as average relative abundance ± standard deviation of three extractions.