Title:

Isolation of chemically well-defined semipreparative liquid chromatography fractions from complex mixtures of proanthocyanidin oligomers and polymers

Authors:

Milla M. Leppä[a], Maarit Karonen[a], Petri Tähtinen[a], Marica T. Engström[a], Juha-Pekka Salminen[a]

Affiliations:

[a]Natural Chemistry Research Group, Department of Chemistry, University of Turku, Vatselankatu 2, FI-20014 Turku, Finland

Corresponding author contact information:

e-mail: mimale@utu.fi

mobile phone number: +358 29 450 3146

postal address: Vatselankatu 2, FI-20014 Turku, Finland

Co-author contact information:

Maarit Karonen (maarit.karonen@utu.fi)

Petri Tähtinen (petri.tahtinen@utu.fi)

Marica T. Engström (mtengs@utu.fi)

Juha-Pekka Salminen (j-p.salminen@utu.fi)

Abstract:

In this study, a semipreparative liquid chromatography method was developed for the isolation of chemically well-defined proanthocyanidin (PA) oligomer and polymer fractions. The aim was to achieve better separation than traditionally achieved for the PAs with other chromatographic methods. The method was tested with eleven PA rich Sephadex LH-20 fractions, which originated from eleven different plant species. The resulting semipreparative fractions were analyzed by both triple quadrupole and high-resolution mass spectrometry assisted by ultrahigh-performance liquid chromatography (UPLC) separation. The results showed remarkable differences in the procyanidin to prodelphinidin ratio, mean degree of polymerization, and specific oligomeric and polymeric content. However, some of these features indicated consistent patterns between species as the function of UPLC retention time. The developed method enables the production of tens of well-defined fractions of PA oligomers and polymers from the unresolved chromatographic PA hump. Accordingly, this allows researchers to explore the most bioactive parts of the complex PA humps of any plant species, which have not been possible earlier.

Keywords:

Fractionation, HRMS, orbitrap, proanthocyanidin, semipreparative LC, UPLC-MS/MS

1 Introduction

Proanthocyanidins (PAs) are oligomers or polymers of flavan-3-ols (Fig. 1). The most common PA subunits are (epi)catechins with 3', 4'-dihydroxylation and (epi)gallocatechins with 3', 4', 5'-trihydroxylation in the B ring. As parts of the oligomeric or polymeric PA structures, (epi)catechins are called procyanidin (PC) units and (epi)gallocatechins as prodelphinidin (PD) units. These subunits are usually linked to each other via $C_4{-}C_8$ or $C_4{-}C_6$ bonds. PAs are abundant for example in legumes [1,2], which are used as animal feeds. In addition, a plant tissue may contain a diverse mixture of PA oligomers and polymers with varying PC/PD composition and degrees of polymerization [3].

Multiple positive bioactivities, such as antioxidant activity [4] and protein binding capacity [5], have been associated with PA-rich plants. It has been proposed, that PAs could play a major role in the ruminant-related bioactivities, such as antimethanogenic [6] and anthelmintic [7–11] activity. This has increased the interest towards PAs and their potential use as nutraceuticals [7,8], since nematodes of ruminants are not expected to build resistance towards plant PAs as rapidly as currently takes place with synthetic anthelmintic drugs [12,13]. If PAs could at the same time decrease methane emissions produced by enteric fermentation processes [14,15], then they could offer at least part of the solution to both, environmental [6] and parasitic [7–11,16] problems related to ruminant production.

The ruminant-related bioactivities of another tannin subgroup, hydrolysable tannins (HTs), have been studied during recent years by using pure HTs. For example, the different structural features of HTs have been linked to their *in vitro* anthelmintic activities [17]. Also, the degree of polymerization of HTs have been studied by the means of protein affinity [18] and antimethanogenic activity [19]. Such a detailed structural information is not available for PAs, since most of the bioactivity studies have been carried out with complex PA mixtures without the knowledge of defined structures of PAs. Thus the structure-bioactivity relationships are not fully

clear for PAs, but for instance the high content of PD subunits and the high mean degree of polymerization (mDP) have been associated with the high anthelmintic activity of PAs *e.g.* in the larval exsheatment inhibition assay [20,21]. To find out more detailed structure–bioactivity relationships for PAs, either pure compounds or chemically well-characterized mixtures of a few PAs, are needed.

A number of different chromatographic methods have been utilized in the PA isolation and purification, usually resulting in the purification of small oligomers (e.g. from dimers to pentamers) or mixtures of polymers. The further purification of the polymeric hump has been one of the main challenges. A common approach in PA fractionation is to use multistep elution with column chromatography, sometimes preceded by liquid-liquid extraction [22]. Liquid-liquid separation can be further exploited with chromatographic methods such as counter current chromatography (CCC) [23–25], and sequential centrifugal partition chromatography (SCPC) [26]. In these methods, two immiscible liquids are used as the mobile and stationary phases, thus ensuring the full recovery of the sample material during the fractionation. For instance, CCC has been used to isolate PA oligomers up to pentamers [23–25].

Different stationary phases and elution techniques have been utilized in column chromatography as well. For instance, Sephadex LH-20 and Toyopearl HW-50F are commonly used size exclusion chromatographic methods, which can produce large quantities of oligomeric and polymeric mixtures of PAs [27,28]. The Sephadex LH-20 chromatography can be carried out as traditional column chromatography, or the bed of the stationary phase can be placed *e.g.* in a Büchner funnel to achieve more rapid but less efficient purification [29]. For more accurate purification, even two-dimensional LC has been utilized [30,31]. Other LC techniques such as normal phase LC, has been used to separate PA oligomers up to hexamer [32] and semipreparative reversed-phase LC has been used to purify individual dimers [33]. However, we are unaware of chromatographic techniques that would be able to fractionate *e.g*. the polymeric PAs produced by

Sephadex LH-20 chromatography, into finer scale fractions containing only a small number of polymers, instead of the polymeric hump.

In this paper, we report a semipreparative liquid chromatography method that enables the controlled production of tens of chemically and chromatographically well-resolved PA fractions from the Sephadex LH-20 fractions rich in PA oligomers and polymers. The semipreparative method is accompanied by fraction analysis with UPLC-MS/MS [34] for the PC/PD ratio and mDP, and high-resolution mass spectrometry (HRMS) analysis for the accurate molecular masses of the PA oligomers and polymers. The method combination was tested with eleven model plant species to show that it was able to produce hundreds of chromatographically narrow PA fractions that varied in their retention time, PC/PD ratio and mDP. These methods provide tools to significantly improve our understanding of the complex nature of plant PAs and their bioactivities by large-scale production of different types of PA isolates followed by accurate structure–activity studies.

2 Materials and methods

2.1 Solvents and materials

Analytical grade acetone and methanol (VWR International S.A.S., France) were used in the extraction and Sephadex LH-20 column cromatography. Semipreparative LC analyses were carried out with HPLC grade acetonitrile (VWR International S.A.S., EC). Formic acid (99−100%, VWR Chemicals, EC) was used on eluent buffering. For the UPLC analyses LC-MS grade acetonitrile was purchased from VWR International S.A.S. (USA) and LC-MS grade formic acid from Sigma Aldrich. The water used was purified with Millipore Synergy UV (Merck KGaA, Darmstadt, Germany) system.

2.2 Sample collection, extraction and Sephadex-LH 20 column chromatography

Different types of tissues of eleven PA rich plant species were selected on the basis of their PA fingerprints to maximize the structural variability (PC/PD ratio, and mDP) of PAs used in this study (Table 1). The plant material was freshly collected into 1L glass bottles, which were then filled with

117 acetone. The samples were macerated in 4 °C for 9−12 months. After the maceration, the first
118 extracts were separated from the plant material and plant tissues were ground. A new batch of
119 extraction solvent (4/1 acetone/water, v/v) was added and the sample materials were dispersed with
120 IKA disperser. Samples were further extracted 5–7 times with 4/1 acetone/water (v/v) until the
121 extraction solvents were colorless. The acetone was evaporated from the solutions and the
122 remaining aqueous extracts were freeze-dried.

123 Approximately 8–10 g of lyophilized extracts were dissolved in 40−50 mL of water. Water
124 insoluble impurities were removed by centrifugation and filtration with 0.45 μm PTFE filters. The
125 dissolved extracts were applied on top of a Sephadex LH-20 column (40 × 4.0 cm Chromaflex®,
126 Kontes). Water phase was eluted and collected in two fractions (Sep1a:1 × 150 mL, Sep1b: 1 × 350
127 mL water), followed by two 500–600 mL methanol/water fractions (Sep2: 3/7 (v/v) MeOH/water,
128 Sep3: 1/1 (v/v) MeOH/water) and three 500–600 mL acetone/water fractions (Sep4: 3/7 (v/v)
129 Me_2CO/water, Sep5: 1/1 (v/v) Me_2CO/water, Sep6: 4/1 (v/v) Me_2CO/water). Finally, the organic
130 solvents were evaporated from the fractions by rotary evaporator and the aqueous fractions were
131 lyophilized. The masses of the freeze-dried PA-rich Sephadex fractions of eleven plant species
132 varied between 270–1630 mg.

133 **2.3 Semipreparative LC**

134 The semipreparative LC samples were prepared by dissolving 125−150 mg of isolated PA-rich Sep6
135 fraction in 3 mL of water. Approximately five drops of ethanol were used to improve the
136 dissolution before the addition of water. The samples were filtrated with 0.2 μm PTFE filters prior
137 to the analysis. An HPLC system with a semipreparative column (150 × 21.20 mm, Gemini®
138 10 μm, C-18, 110Å, Axia packed, Phenomenex) was used for further isolation of the studied PAs.
139 The HPLC system consisted of a Waters 2535 Quaternary Gradient Module (Water Corp., USA), a
140 Waters 2998 photodiode array Detector (Waters Corp., Singapore), and a Waters Fraction Collector
141 III (Waters Corp., Japan). Acetonitrile (A) and 0.1% aqueous formic acid (B) were used as eluents.

For the elution, 12.0 mL min^{-1} flow rate was used and fractions were collected into 2 ml tubes from 5 to 33 min resulting in total of 168 individual fractions per one fractionation and each fraction containing the PAs eluting during 10 s retention time window. Before the elution, the column was thoroughly stabilized for 30 mins with 8% A in B. The elution was performed as follows: 0−4 min, isocratic 8% A in B; 4−32 min, 8−55% A in B (linear gradient); 32−35 min, 55−80% A in B (linear gradient), 35−80 min column wash and stabilization. From the retention time area, which inheld the complete PA fingerprint as detected by UV detection at 280 nm, every fourth fraction was selected and analyzed by UPLC-DAD-MS/MS and UPLC-DAD-HRMS.

2.4 UPLC-DAD-MS/MS analysis

PC and PD units were detected with specific single reaction monitoring (SRM) methods for both extension and terminal units to allow the calculation of both PC and PD concentrations and mDP [34]. The detection of monomers, oligomers and polymers was enabled by the use of four different cone voltages in the fragmentation of the PC and PD units from the original PAs. Tandem mass analyses were performed with an Aquity UPLC system (Waters Corp., Milford, MA, USA) coupled to a Xevo TQ triple quadrupole mass spectrometer (Waters Corp., Milford, MA, USA). The UPLC system consisted of a sample manager, a binary solvent manager, a column, and a diode array detector. The column used was a Waters Acquity UPLC BEH Phenyl (1.7 μm, 2.1 × 100 mm Waters Corp. Wexferd, Ireland). For the elution, acetonitrile (A) and 0.1% aqueous formic acid (B) were used with 0.5 mL min^{-1} flow rate. The elution profile was performed as follows: 0−0.5 min, 0.1% A in B (isocratic); 0.5−5.0, 0.1−30% A in B (linear gradient); 5.0−8.0 min, 30−45% A in B (linear gradient); 8.0−11.5 min, column wash and stabilization. The data was recorded from 0 to 8 min.

Negative ion mode was used in MS analyses, with electrospray ionization conditions described by Engström [34]. In short, the following values were used: capillary voltage 2.4 kV,

desolvation and source temperature 650 °C and 150 °C, respectively, desolvation and cone gas (N_2) flow 100 and 1000 L h^{-1},respectively. The stability of ionization was monitored with 1 μg mL^{-1} catechin solution (in acetonitrile/0.1% formic acid, v/v), which was analyzed five times before and after each batch of 10 samples [1]. The quantitation of PC and PD units and determination of mDP was performed as described by Engström et al. [34] and Malisch et al. [1]. The recorded PC and PD traces were smoothed (window size 5 scans × 2 smoothing iterations) with the Target Lynx software (V4.1 SCN876 SCN 917 © 2012 Waters Inc.) and converted into quantitative data by the help of calibration curves made separately for PC, PD and mDP. The PC and PD calibration curves were prepared by making dilutions in known concentrations (1.50–0.1875 mg mL^{-1} for the PC standard and 2.00–0.25 mg mL^{-1} for the PD standard) of two Sephadex LH-20 fractions with known PC and PD content that was initially calibrated against thiolysis results [34]. Calibration curves for the mDP were similarly obtained with six Sephadex LH-20 fractions with known PC and PD content and mDP.

2.5 UPLC-DAD-HRMS analysis

The analyses for accurate masses were carried out using an Aquity UPLC system (Waters Corp., Milford, MA, USA) which was coupled to a quadrupole–Orbitrap mass spectrometer (Q ExactiveTM, Thermo Fisher Scientific GmbH, Bremen, Germany). The UPLC system was identical to the previously mentioned UPLC system, apart from the column, which in this case was an Aquity UPLC BEH Phenyl 1.7 μm, 2.1 × 30 mm column (Waters, Ireland). A flow rate of 0.65 mL min^{-1} was utilized with the same solvents as described in the previous chapter. The elution profile was as follows: 0−0.1 min, 3% A in B (isocratic); 0.1−3.0 min, 3−45% A in B (linear gradient); 3.0−4.2 min, column wash and stabilization. The UV (λ = 190–500 nm) and MS data was detected throughout the analysis and analytes were detected as negative ions. A heated ESI source (H-ESI II, Thermo Fisher Scientific GmbH) was utilized with similar ionization source parameters as described by Suvanto *et al* [35]. In brief, the following parameters were used in the ionization

192 source: spray voltage, −3.0 kV; capillary temperature, 380 °C; sheat, aux and sweep gas (N_2) flow
193 rate, 60, 20 and 0 arbitary units respectively. In the orbitrap detector, the mass range was set at
194 *m/z* 200–3000, the resolution was set at 70,000, and the automatic gain of 3×10^6 was used.
195
196 **3 Results and discussion**
197 **3.1 Sephadex LH-20 fractionation of eleven plant species**
198 Plant extracts were fractionated with Sephadex LH-20 column chromatography to produce
199 PA-rich fractions. Since Sephadex LH-20 chromatography of tannins is largely based on their
200 affinity to the gel rather than size exclusion, smaller PAs typically elute from the column before the
201 large ones and the largest tannins are efficiently retained in the gel [36]. Thus, PAs eluted mainly in
202 the two last fractions, Sep5 and Sep6. Sep5 fractions contained greater amount of other
203 polyphenols, such as flavonoids (Supplementary material, Fig. 1) while Sep6 fractions contained
204 mainly PA oligomers and polymers. Thus, Sep6 fractions were selected as the PA sources for the
205 further purification by semipreparative HPLC. The UPLC-DAD chromatograms (λ = 280 nm),
206 PC/PD ratios and mDPs of the Sep6 fractions are displayed in supplementary material
207 (Supplementary material, Fig. 2). Considerable variation was observed in the widths of the
208 polymeric PA humps by UPLC-DAD. PAs of the most PD-rich fractions started to elute at the
209 retention time of 2 min, while the PAs in the most PC-rich samples started to elute at 3 min or later.
210 Since the PC-% and mDP varied between 2–99% and 4–20, respectively, these eleven fractions
211 served together as a good source of PAs with respect to structural diversity for testing the efficiency
212 of the semipreparative LC method.

213 **3.2 Semipreparative LC fractionation of the Sep6 fractions**
214 Semipreparative LC fractionation of Sep6 fractions (Fig. 2) showed similar chromatographic
215 profiles than was earlier obtained by UPLC. In semipreparative HPLC conditions, the elution of the
216 PA humps started at 7–15 mins and lasted until at 22–30 mins, depending on the Sep6 fraction.

Every fourth semipreparative LC fraction (Sem fraction) was then analyzed by UPLC-DAD-MS/MS for their PC/PD ratio, mDP and UV chromatogram. Figure 3 shows the UV chromatograms (λ = 280 nm) of *Lysimachia vulgaris* Sem fractions as an example to highlight how the semipreparative approach allowed the production of chromatographically defined PA fractions. In the UPLC analyses, the UV peak widths of Sem fractions varied between 0.3−1.2 min (Fig. 3) while the peak widths of original Sep6 fractions varied between 2.0–4.5 min (Supplementary Fig. 2). This clearly demonstrates the efficiency of the semipreparative LC fractionation in the separation of PAs.

In general, the isolated Sem fractions preserve their elution order (Fig. 3, Supplementary material Fig. 3A) when comparing the two chromatographic techniques. However, some irregularities were observed in the UPLC chromatograms of the Sem fractions. For instance, the latest Sem fractions produced wide UPLC peaks at earlier retention times than the previous fractions (Supplementary material Fig. 3B). This was interpreted to be due to the reversed-phase LC columns being unable to separate properly the largest PAs [37]. In addition, for instance Sem fractions 35, 39, 107 and 115 (Fig. 3), showed sharp UPLC peaks on the top of the more typical PA hump. These peaks were due to small PA oligomers or non-tannin impurities, especially flavonoids that elute simultaneously with the PA hump.

3.3 UPLC-DAD-MS/MS results of the semipreparative LC fractions

The UPLC-DAD-MS/MS [34] analyses revealed the PC and PD traces of the Sem fractions. Figure 4 shows the PA fingerprints of 20 selected Sem fractions with different PC/PD ratios, mDPs and retention times. These examples revealed how the used fractionation method enabled the production of a highly diverse selection of different PA fractions ranging from almost PC pure to almost PD pure. For instance, the almost PC pure fractions (98–99% PC), which eluted from 3.5 to 6 minutes (Fig. 4A−D), were different either by their mDP (4–17) or retention time. Thus, they were all chemically different. For example, *Aesculus hippocastanum* Sem fraction 116 (Fig. 4C) and

242 *Trifolium medium* Sem fraction 82 (Fig. 4B) had similar PC/PD ratios and mDP, but they still
243 eluted at different retention times. One explanation for this may lay in the composition of PA
244 oligomers and polymers. *A. hippocastanum* Sem fractions (Fig. 4C) were rich in A-type PAs whilst
245 only B-type PAs were found in the *T. medium* Sem fractions (Fig. 4B). Similar variability was
246 observed with different PC/PD mixtures (Fig. 4E−R) and almost pure PDs (Fig. 4S−T). For
247 instance, *Lotus corniculatus* Sem fraction 92 (Fig. 4J) and *Pinus sylvestris* Sem fraction 78
248 (Fig. 4K) had almost similar mDPs and PC/PD ratios, but they differed from each other by retention
249 times which were 4.5 and 4.0 min respectively.

250 Altogether, hundreds of fractions were obtained with either different PC/PD ratios, mDPs
251 and/or retention times. The retention time of the fraction as such did not reveal much of their
252 chemical composition. A fraction eluting at 4 mins could be either 99% PC with the mDP of 4 (Fig.
253 4A), 60% PC with the mDP of 18 (Fig. 4I), or 30% PC with the mDP of 30 (Fig. 4P). Such a
254 diversity of PA chemistry within the hundreds of fractions, enables the more accurate determination
255 of structure–activity relationships of PAs. The above-mentioned data of the hundreds of
256 semipreparative fractions were used to create species-specific elution profiles of PC/PD ratios and
257 mDPs.

258 **3.4 Distribution of PCs, PDs and mDP in semipreparative LC elution**

259 The distribution of PC and PD units during the semipreparative LC analysis varied significantly
260 among different samples. To make the conclusions more uniform, the samples were categorized
261 into five groups depending on their total PC/PD ratio **I**: PC-% ≥ 98%; **II**: PC-% = 97–61%; **III**: PC-
262 % = 60–41 %; **IV**: PC-% = 40–21%; **V**: PC-% ≤ 20% (group III included only one sample,
263 therefore the figure is displayed in supplementary material only, Supplementary Fig.4).

264 Group **I** included Sep6 fractions of *Trifolium medium* and *Aesculus hippocastanum*. In the
265 group **I,** the PC content was constant at 99% (Fig. 5A) for almost throughout the analysis. Only the
266 first four analyzed fractions from the retention time of 12–15 mins contained 1–17% PD units. In

total, the mDP increased during the elution, but with *A. hippocastanum* (supplementary data Fig. 4) the mDP decreased slightly in at certain retention time points. The decreases appear at the same retention times as single peaks in the semipreparative UV chromatogram (λ = 280 nm). Thus, the early eluting smaller oligomers temporarily distorted the mDP curve.

Group **II** included *Larix sp.*, *Rhododendron schlippenbachii* Maxim. and *Rhododendron dichroanthum* Diels. In group **II** (Fig. 5B and supplementary Fig. 4), an interesting pattern in PC/PD ratio was discovered. When the PAs started to elute, the first eluting compounds were PD-rich oligomers. However, at the retention time of 13.2−15.4 min, the most abundant subunit type changed from PD to PC, after which, the PC content was settled between 60−90%. Apart from the other groups, the mDP did not increase continuously during elution in group **II**.

Group **III** contained one sample, the Sep6 fraction of *Lotus corniculatus* (Supplementary Fig. 4). Nevertheless, a characteristic feature for this sample was that, the mDP increased during the elution and the PC/PD ratio settled close to 50/50 after 15.6 min retention time. The results indicated that the main compounds in the *L. corniculatus* were PAs consisting of both subunits, PC and PD rather than PC or PD pure oligomers.

Group **IV** included the Sep6 fractions of *Lysimachia vulgaris* and *Pinus sylvestris*. In the **IV** group, the mDP increased relatively steadily throughout the elution while the PC/PD ratio varied greatly. For example the PD content of the *L. vulgaris* (Fig. 5C) sample, varied between 97–31% during the elution. An interesting pattern of PC/PD ratio was observed with *Pinus sylvestris* fractions as well (Supplementary Fig. 5). At the beginning of the elution, the PD content was > 80%, after which it decreased to 55%. But after the retention time of 16.5−18.6 min, the amount of PD units increased again up to 73%. PD units are considered more hydrophilic than the PC units, thus according to a general assumption, the PD-rich PAs should elute prior to the PC-rich ones. This data however indicated that the PD content could increase during the elution. Also, the mDP increased during the elution. Hence, the late eluting polymers had two structural features, which

affect the retention, the early elution due to additional -OH groups and the late elution due to the large polymer size [1].

The last group included the Sep6 fractions of *Salix phylicifolia*, *Ribes alpinum* and *Trifolium repens*. In the group **V** (Fig. 5D, Supplementary Fig. 5), the mDP shows an unusual pattern. In general, the mDP increased during the elution, but at 3–5 min after the most intensive point at semipreparative chromatogram (λ = 280 nm), the mDP reached a local maximum, after which the mDP decreased again. This indicates that the largest oligomeric and polymeric PAs in PD-rich samples eluted 3–5 min after the most abundant and slightly smaller PAs. PDs remained as the most abundant subunit type throughout the analysis in the group **V**.

Even though the samples were diverse in their PC/PD ratios and mDP distributions, some characteristic features were common for all analyzed samples. For example, the PD-% values were highest at the beginning of the elution in all five groups. The increased hydrophilicity of the PD rich oligomers was the most likely cause of the early elution. In addition, the changes of mDP during the elution were variable between different PA composition groups. The most usual trend of the mDP was, that it increased by retention time. The increase was most considerable in PC (group **I**) and PD (group **V**) rich groups whilst in the group **II** the changes in mDP did not seem to have a systematic pattern. It seemed that, when the used Sep6 fractions were either PC or PD rich, the mDP increased steadily during the elution. In these cases, the main factor affecting the retention was the size of the oligomers, thus the largest oligomers eluted later.

Altogether, the PA mixtures included a great variety of different mDPs and subunit compositions. For example, the Sep6 fraction of *Larix sp.* (Fig. 5B) had the average PC/PD ratio of 62/38 whereas the most PD-rich fraction from semipreparative LC analysis contained 97% PDs and the most PC-rich fraction contained only 11% PDs. Only few fractions had the average PC/PD ratio of approximately 60/40 and generally, the composition of the produced semipreparative LC fractions differed significantly from the average composition of the Sep6 fractions. Thus, it is

possible to isolate a large variety of different PA mixtures from a single pre-purified plant sample with the semipreparative LC elution protocol introduced in this paper.

3.5 UPLC-DAD-HRMS results of the semipreparative LC fractions

All semipreparative fractions that were analyzed by quantitative MS/MS were analyzed also by high-resolution mass spectrometry (HRMS) with an UPLC-Q-Orbitrap-MS instrument. PAs are prone to fragmentation [38,39] thus the possibility of detecting quinone-methide fragments as molecular or multiply charged ions was taken into account at the early stage of the mass spectral interpretation. Figure 6 displays the retention time shifts of multiple extracted ion chromatograms (EICs) of PC-rich PAs in *Lysimachia vulgaris* Sem fraction 91. The chromatographic humps of PAs shifted to latter retention times when the DP (degree of polymerization) of oligomers and polymers increased. EICs overlapped, but since they were not identical, they probably originated from different compounds. Thus, the most significant peaks in the forecoming high-resolution mass spectra, were interpreted as molecular $[M–H]^–$ or multiply charged ions $[M–XH]^{X–}$ instead of fragments. Herein, the observations from high-resolution mass data are described through four representative semipreparative LC fractions of *Lysimachia vulgaris* and *Larix sp*.

Large proanthocyanidins tend to form multiply charged ions in the electrospray ionization source [40]. Depending on the degree of polymerization, the PAs formed doubly (DP 4–11, Fig. 7 and 9), and triply charged ions (DP 10–20, Fig. 7 and 8). Even four-fold charged ions were observed (DP > 18) in the *L. vulgaris* Sem fraction 91 (Fig. 8). Due to the multiply charged ions, certain patterns were observed in the mass spectra. In the case of PC-rich samples, the PC pure oligomers and polymers were observed as at *m/z* 577.13 + 288.06n for singly [37,41], at *m/z* 576.13 + 144.03n for doubly and at *m/z* 959.87 + 96.02n for triply charged ions (Fig. 8). In the case of PD-rich samples the PA oligomers and polymers were observed as at *m/z* 609.12 + 304.06n for singly, at *m/z* 608.12 + 152.03n for doubly and at *m/z* 1013.19 + 101.35n for triply charged ions. The smallest observed oligomers for each charge state were dimers (z = 1–), tetramers (z = 2–) and

342 decamers (z = 3–). The number of additional PC or PD units is described with n, where n ≥ 0 is the
343 amount of additional PC or PD unit to the smallest oligomers.

344 Systematic oligomer series were observed within each polymer size as well. Within a single
345 oligomer size, an increment of *m/z* +16.00 for singly, +8.00 for doubly, +5.33 for triply and +4.00
346 for fourfold charged ions was observed. These systematic increases in the mass spectra were caused
347 by a difference of one subunit being PD instead of PC. For instance, *L. vulgaris* Sem fraction 67
348 contained a doubly charged PC pure heptamer at *m/z* 1008.22 and a doubly charged PC rich
349 heptamer consisting of 6 × PC + 1 × PD unit at *m/z* 1016.22 (Fig. 7C–D and 8C). Thus, their *m/z*
350 difference was 8.00.

351 The differences between *L. vulgaris* Sem fraction 67 (Fig. 7) and 91 (Fig. 8) demonstrated the
352 changes of PA oligomer composition in different retention time points of the semipreparative
353 elution. The Sem fraction 67 of *L. vulgaris* included a mixture of oligomers ranging from PC pure
354 to PD pure (Fig. 7, Table 2) and the degree of polymerization ranged between 4–17 subunits.
355 Within at *m/*z range of 960–1120 (Fig. 7B), three oligomeric and polymeric PA series were
356 observed. These were doubly charged series of heptameric (DP 7) and triply charged series of
357 decameric (DP 10) and undecameric (DP 11) PAs. In the case of heptrameric oligomer series (Fig.
358 7C), the PC pure oligomer (at *m/z* 1008.2207, Table 2) was observed as the most intensive
359 oligomer. The oligomeric composition of heptamer ranged from 7 × PC + 0 × PD units (at
360 *m/z* 1008.2207) to 1 × PC + 6 × PD units (at *m/z* 1056.2028). The *m/z* addition to the previous
361 signal was 8.00 Da and all PC/PD combinations were found except for PD pure heptamer. In the
362 undecameric PA series (Fig. 7D), the most abundant oligomer consisted of 3 × PC + 8 × PD units
363 (at *m/z* 1098.5461, Table 2). The undecameric oligomer series ranged from 10 × PC + 1 × PD units
364 (at *m/z* 1061.2260) to 1× PC + 10 × PD units (at *m/z* 1109.2085). The *m/z* addition to the previous
365 signal was 5.33 Da, when the oligomers differed from each other by one subunit. Similar
366 distribution patterns of PC and PD units were observed with other oligomer sizes as well, thus

indicating that the *L. vulgaris* Sem fraction 67 was a diverse mixture of oligomeric and polymeric PAs containing both PC and PD subunits.

Sem fraction 91 of *L. vulgaris* consisted mainly of PC-rich oligomers (Fig. 8A, Table 3) and the most intensive ions in the mass spectrum were PC pure oligomers. The polymerization of the compounds ranged between 5–25 subunits. The closer examination of the spectrum revealed the composition of subunits in more detail. For example at the *m/z* range of 1510–1560 (Fig. 8B), polymers consisting of 16 (z = 3–) and 21 (z = 4–) subunits were observed. The most intensive compounds in both PA series were PC-pure polymers. In the fourfold charged 21-meric PA series, the polymers ranged from 21 × PC + 0 × PD (at *m/z* 1511.8209, Fig. 8C, Table 3) to 17 × PC + 4 × PD units (at *m/z* 1527.8123). In the hexadecameric series of PAs, the polymers ranged from 16 × PC + 0 × PD (at *m/z* 1535.9923, Table 3) to 14 × PC + 2 × PD units (at *m/z* 1546.6559). In the polymeric series of the fourfold charged ions, the addition of 4.00 Da was observed to the previous signal, when the PC/PD subunit content increased by one PD unit.

The high-resolution mass spectra showed significantly differing PA compositions in *L. vulgaris* Sem fractions 67 and 91. The qualitative differences between the high-resolution mass spectra, were similar to the differences observed in the qualitative MS/MS data. Larger polymers were identified in the Sem fraction 91 (mDP = 27) than in the Sem fraction 67 (mDP = 23). Also, Sem fraction 91 (PC/PD = 68/32) was distinctively PC-rich whereas Sem fraction 67 (PC/PD = 36/64) consisted of both PC and PD-rich oligomers and polymers. The difference in the ionization efficiency of PC and PD units is the most probable reason why Sem fraction 91 seemed to contain such PC pure composition whilst Sem fraction 67 consisted of a large variety of PC and PD containing oligomers. PC-rich oligomers are more easily ionized, and therefore, their relative abundance in the mass spectra might seem excessive. Nevertheless, the diversity of the PA oligomers and polymers was more significant with Sem fraction 67 than in Sem fraction 91. This

was observed especially from the numbers of different PC/PD unit combinations in each oligomeric or polymeric PA series.

Another representative example of the differences which emerge during the semipreparative LC, were observed in *Larix sp.* Sem fractions 74 and 94 (Fig. 9). Both of these samples were PC-rich and the most intensive signals of the oligomers were PC pure oligomers. The most considerable differences between the mass spectra of fractions 74 and 94 were found in the relative intensity of individual oligomers and their polymeric size. The largest polymer in *Larix* sp. Sem fraction 74 was 14 × PC (Fig. 9A), and in Sem fraction 94 it was 20 × PC (Fig. 9B). In the fraction 74, the most abundant PA oligomers ranged from dimer to hexamer, whereas the fraction 94 included a variety of oligomers with relative abundance of > 40 %, ranging from 6 to 13 subunits. These differences in high-resolution mass spectra were similar to the differences observed in the mDP measured by tandem MS. The fraction 74 contained a less complex composition of PAs and the most abundant oligomers were smaller than in the fraction 94. The mDPs of the above-mentioned fractions were 8.4 and 13.3 respectively.

The separated fractions were highly complex in PA composition, thus, the original mixtures of PA oligomers and polymers must have been even more complex. The fractionation of PA oligomers enabled the detailed identification of PAs by high-resolution mass spectra. By fractionation, even the PAs in minor concentrations were accumulated and identified. These findings show that the average PC/PD or mDP values of the PA mixtures produced by e.g. Sephadex LH-20 gel chromatography, did not give a rightful impression of the individual compounds in the samples. Instead, individual PA isolates produced from the Sephadex LH-20 fractions, may possess such different composition of oligomers and polymers that cannot be estimated from the mDP or average PC/PD ratio of the original plant sample or Sephadex LH-20 fraction.

4 Conclusions

In present study, we developed a semipreparative LC method, which can be exploited in the fractionation of pre-purified PAs into tens of different PA fractions. The retention time windows of the fractionated PAs were significantly narrower than the retention time windows of the original PA samples. The semipreparative fractions also preserved their elution order in the UPLC conditions. The closer investigation of the semipreparative PA fractions revealed a great variability in PC/PD compositions, mDPs and retention times. Semipreparative elution profiles of the PC/PD ratios and mDPs were presented along with the semipreparative LC chromatograms (λ = 280 nm). The elution patterns of PC and PD units and mDPs were rather species specific and highly dependent on the PA composition of the Sephadex-LH 20 fractions. The main conclusions regarding the semipreparative elution were the early elution of PD rich PAs compared to the PC rich ones, and the elution of smaller oligomers prior to larger ones in the case of either PC or PD rich sample. Altogether, major changes in PC/PD ratios emerged during semipreparative elution, thus indicating high versatility of PA molecules in the Sephadex-LH 20 fractions.

Individual PAs were identified with high-resolution mass spectrometry. The identification of compounds confirmed the quantitative MS/MS measurements and revealed the PC/PD variability within single fractions. Even fractions purified from the very same plant source, were divergent in their subunit composition and polymer size. This study elaborates our understanding about the PA composition beneath the previously unresolved chromatographic hump and provides tools for the production of tens and hundreds of PA isolates with known chemical composition.

5 Appendix 1

DAD, diode array detector; DP, degree of polymerization; mDP, mean degree of polymerization; PA, proanthocyanidin; PC, procyanidin; PC/PD, the ratio of procyanidin and prodelphinidin units; PD, prodelphinidin; QqQ, triple quadrupole; Sep5; Sephadex LH-20 fraction 5, Sep6, Sephadex

LH-20 fraction 6; SRM, single reaction monitoring; UPLC, ultrahigh-performance liquid chromatography.

6 Acknowledgements

We thank Anne Koivuniemi for the maintenance of the analysis equipment and Suvi Vanhakylä for the collection of the plant material. The crew of Natural Chemistry Research Group is acknowledged for all the general help in the laboratory.

7 Funding Sources

We acknowledge Raisio Research Foundation and Academy of Finland (298177) for the funding of this research. A purchase of the UHPLC-DAD-QqQ system was made possible by a Strategic Research Grant of the University of Turku (Ecological Interactions).

8 Supplementary Material Description

Figure 1. UPLC-DAD UV traces (λ = 280 nm) of Sephadex LH-20 fractionation of *Larix sp.*

Figure 2. UPLC–DAD UV traces (λ = 280 nm) of Sep6 fractions of the eleven plant species. The abbreviations are as follows: Sep6 fraction, the 6th Sephadex LH-20 fraction eluted with 4/1 (v/v) acetone water; mDP, mean degree of polymerization; PC/PD, ratio of procyanidin and prodelphinidin units.

Figure 3. UPLC-DAD UV traces (λ = 280 nm) of semipreparative LC fractionation of *Ribes alpinum* and *Trifolium medium*.

Figure 4. Semipreparative LC DAD traces (λ = 280 nm) with mDP and PC/PD distribution of *Aesculus hippocastanum*, *Rhododendron dichroanthum*, *Rhododendron schlippenbachii* and *Lotus corniculatus.*

Figure 5. Semipreparative LC DAD traces (λ = 280 nm) with mDP and PC/PD distribution of *Pinus sylvestris*, *Salix phylicifolia* and *Trifolium repens*.

9 References

[1] C.S. Malisch, A. Lüscher, N. Baert, M.T. Engström, B. Studer, C. Fryganas, D. Suter, I. Mueller-Harvey, J.P. Salminen, Large Variability of Proanthocyanidin Content and Composition in Sainfoin (*Onobrychis viciifolia*), J. Agric. Food Chem. 63 (2015) 10234–10242. doi:10.1021/acs.jafc.5b04946.

[2] J.H. Niezen, T.S. Waghorn, W.A.G. Charleston, G.C. Waghorn, Growth and gastrointestinal nematode parasitism in lambs grazing either lucerne (*Medicago sativa*) or sulla (*Hedysarum coronarium*) which contains condensed tannins, J. Agric. Sci. 125 (1995) 281. doi:10.1017/S0021859600084422.

[3] I. Mueller-Harvey, G. Bee, F. Dohme-Meier, H. Hoste, M. Karonen, R. Kölliker, A. Luscher, V. Nidekorn, W.F. Pellikaan, J.-P. Salminen, L. Skot, L.M.J. Smith, S.M. Thamsborg, P. Totterdell, I. Wilkison, A.R. Williams, B.N. Azuhnwi, N. Baert, A.G. Brinkhaus, G. Copani, O. Desrues, M.T. Engström, C. Fryganas, M. Girard, N.T. Huyen, K. Kempf, C. Malisch, M. Mora-Ortiz, J. Quijada, A. Ramsay, H.M. Ropiak, G.C. Waghorn, Benefits of Condensed Tannins in Forage Legumes Fed to Ruminants: Importance of Structure, Concentration and Diet Composition, Crop Sci. (2017). doi:10.2135/cropsci2017.06.0369.

[4] Y. Gong, F. Fang, X. Zhang, B. Liu, H. Luo, Z. Li, B Type and Complex A/B Type Epicatechin Trimers Isolated from *Litchi pericarp* Aqueous Extract Show High Antioxidant and Anticancer Activity, Int. J. Mol. Sci. 19 (2018) doi:10.3390/ijms19010301.

[5] W.E. Zeller, M.L. Sullivan, I. Mueller-Harvey, J.H. Grabber, A. Ramsay, C. Drake, R.H. Brown, Protein precipitation behavior of condensed tannins from *Lotus pedunculatus* and *Trifolium repens* with different mean degrees of polymerization, J. Agric. Food Chem. 63

(2015) 1160–1168. doi:10.1021/jf504715p.

[6] B. Hatew, E. Stringano, I. Mueller-Harvey, W.H. Hendriks, C.H. Carbonero, L.M.J. Smith, W.F. Pellikaan, Impact of variation in structure of condensed tannins from sainfoin (*Onobrychis viciifolia*) on *in vitro* ruminal methane production and fermentation characteristics, J. Anim. Physiol. Anim. Nutr. (Berl). 100 (2016) 348–360. doi:10.1111/jpn.12336.

[7] H. Hoste, J.F.J. Torres-Acosta, C.A. Sandoval-Castro, I. Mueller-Harvey, S. Sotiraki, H. Louvandini, S.M. Thamsborg, T.H. Terrill, Tannin containing legumes as a model for nutraceuticals against digestive parasites in livestock, Vet. Parasitol. 212 (2015) 5–17. doi:10.1016/j.vetpar.2015.06.026.

[8] H. Hoste, F. Jackson, S. Athanasiadou, S.M. Thamsborg, S.O. Hoskin, The effects of tannin-rich plants on parasitic nematodes in ruminants, Trends Parasitol. 22 (2006) 253–261. doi:10.1016/j.pt.2006.04.004.

[9] J. Quijada, C. Fryganas, H.M. Ropiak, A. Ramsay, I. Mueller-Harvey, H. Hoste, Anthelmintic Activities against *Haemonchus contortus* or *Trichostrongylus colubriformis* from Small Ruminants Are Influenced by Structural Features of Condensed Tannins, J. Agric. Food Chem. 63 (2015) 6346–6354. doi:10.1021/acs.jafc.5b00831.

[10] A.L. Molan, Effect of purified condensed tannins from pine bark on larval motility, egg hatching and larval development of *Teladorsagia circumcincta* and *Trichostrongylus colubriformis* (Nematoda: Trichostrongylidae), Folia Parasitol. (Praha). 61 (2014) 371–376. doi:10.14411/fp.2014.036.

[11] A.R. Williams, H.M. Ropiak, C. Fryganas, O. Desrues, I. Mueller-Harvey, S.M. Thamsborg, Assessment of the anthelmintic activity of medicinal plant extracts and purified condensed tannins against free-living and parasitic stages of *Oesophagostomum dentatum*, Parasit. Vectors. 7 (2014) 518. doi:10.1186/s13071-014-0518-2.

514 [12] R.M. Kaplan, Drug resistance in nematodes of veterinary importance: A status report, Trends
515 Parasitol. 20 (2004) 477–481. doi:10.1016/j.pt.2004.08.001.
516 [13] J.F.J. Torres-Acosta, P. Mendoza-de-Gives, A.J. Aguilar-Caballero, J.A. Cuellar-Ordaz,
517 Anthelmintic resistance in sheep farms: update of the situation in the American continent,
518 Vet. Parasitol. 189 (2012) 89–96. doi:10.1016/j.vetpar.2012.03.037.
519 [14] J.A. Foley, N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D.
520 Mueller, C. O'Connell, D.K. Ray, P.C. West, C. Balzer, E.M. Bennett, S.R. Carpenter, J.
521 Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, D.P.M. Zaks,
522 Solutions for a cultivated planet, Nature. 478 (2011) 337–342. doi:10.1038/nature10452.
523 [15] F.P. O'Mara, The significance of livestock as a contributor to global greenhouse gas
524 emissions today and in the near future, Anim. Feed Sci. Technol. 166–167 (2011) 7–15.
525 doi:10.1016/j.anifeedsci.2011.04.074.
526 [16] O. Desrues, C. Fryganas, H.M. Ropiak, I. Mueller-Harvey, H.L. Enemark, S.M. Thamsborg,
527 Impact of chemical structure of flavanol monomers and condensed tannins on *in vitro*
528 anthelmintic activity against bovine nematodes., Parasitology. 143 (2016).
529 [17] M.T. Engström, M. Karonen, J.R. Ahern, N. Baert, B. Payré, H. Hoste, J.P. Salminen,
530 Chemical Structures of Plant Hydrolyzable Tannins Reveal Their in Vitro Activity against
531 Egg Hatching and Motility of *Haemonchus contortus* Nematodes, J. Agric. Food Chem. 64
532 (2016) 840–851. doi:10.1021/acs.jafc.5b05691.
533 [18] M. Karonen, M. Oraviita, I. Mueller-Harvey, J.P. Salminen, R.J. Green, Binding of an
534 Oligomeric Ellagitannin Series to Bovine Serum Albumin (BSA): Analysis by Isothermal
535 Titration Calorimetry (ITC), J. Agric. Food Chem. 63 (2015) 10647–10654.
536 doi:10.1021/acs.jafc.5b04843.
537 [19] N. Baert, W.F. Pellikaan, M. Karonen, J.-P. Salminen, A study of the structure-activity
538 relationship of oligomeric ellagitannins on ruminal fermentation *in vitro*, J. Dairy Sci. 99

539 (2016) 8041–8052. doi:10.3168/jds.2016-11069.
540 [20] J. Quijada, C. Fryganas, H.M. Ropiak, A. Ramsay, I. Mueller-Harvey, H. Hoste,
541 Anthelmintic Activities against *Haemonchus contortus* or *Trichostrongylus colubriformis*
542 from Small Ruminants Are Influenced by Structural Features of Condensed Tannins, J.
543 Agric. Food Chem. 63 (2015) 6346–6354. doi:10.1021/acs.jafc.5b00831.
544 [21] V. Spiegler, E. Liebau, A. Hensel, Medicinal plant extracts and plant-derived polyphenols
545 with anthelmintic activity against intestinal nematodes, Nat. Prod. Rep. 34 (2017) 627–643.
546 doi:10.1039/C6NP00126B.
547 [22] T.J. Bond, J.R. Lewis, A. Davis, A.P. Davis, Analysis and Purification of Catechin and Their
548 Transformation Products, in Methods in Polyphenol Analysis, C. Santos-Buelga, G.
549 Williamson (Eds.), (2003) pp. 229–265.
550 [23] T. Esatbeyoglu, V. Wray, P. Winterhalter, Isolation of dimeric, trimeric, tetrameric and
551 pentameric procyanidins from unroasted cocoa beans (*Theobroma cacao* L.) using
552 countercurrent chromatography, Food Chem. 179 (2015) 278–289.
553 doi:10.1016/j.foodchem.2015.01.130.
554 [24] T. Esatbeyoglu, P. Winterhalter, Preparation of dimeric procyanidins B1, B2, B5, and B7
555 from a polymeric procyanidin fraction of black chokeberry (*Aronia melanocarpa*), J. Agric.
556 Food Chem. 58 (2010) 5147–5153. doi:10.1021/jf904354n.
557 [25] N. Köhler, V. Wray, P. Winterhalter, Preparative isolation of procyanidins from grape seed
558 extracts by high-speed counter-current chromatography, J. Chromatogr. A. 1177 (2008) 114–
559 125. doi:10.1016/j.chroma.2007.11.028.
560 [26] V. Pedan, N. Fischer, S. Rohn, Extraction of cocoa proanthocyanidins and their fractionation
561 by sequential centrifugal partition chromatography and gel permeation chromatography,
562 Anal. Bioanal. Chem. (2016) 5905–5914. doi:10.1007/s00216-016-9705-7.
563 [27] R.H. Brown, I. Mueller-harvey, W.E. Zeller, L. Reinhardt, E. Stringano, A. Gea, C. Drake,

564 H.M. Ropiak, C. Fryganas, A. Ramsay, E.E. Hardcastle, Facile Purification of Milligram to
565 Gram Quantities of Condensed Tannins According to Mean Degree of Polymerization and
566 Flavan-3-ol Subunit Composition, J. Agric. Food Chem. 65 (2017) 8072–8082.
567 doi:10.1021/acs.jafc.7b03489.
568 [28] Z. Guadalupe, A. Soldevilla, M.P. Sáenz-Navajas, B. Ayestarán, Analysis of polymeric
569 phenolics in red wines using different techniques combined with gel permeation
570 chromatography fractionation, J. Chromatogr. A. 1112 (2006) 112–120.
571 doi:10.1016/j.chroma.2005.11.100.
572 [29] N. Baert, M. Karonen, J.P. Salminen, Isolation, characterisation and quantification of the main
573 oligomeric macrocyclic ellagitannins in *Epilobium angustifolium* by ultra-high performance
574 chromatography with diode array detection and electrospray tandem mass spectrometry, J.
575 Chromatogr. A. 1419 (2015) 26–36. doi:10.1016/j.chroma.2015.09.050.
576 [30] L. Montero, V. Sáez, D. von Baer, A. Cifuentes, M. Herrero, Profiling of *Vitis vinifera* L.
577 canes (poly)phenolic compounds using comprehensive two-dimensional liquid
578 chromatography, J. Chromatogr. A. 1536 (2018) 205–215.
579 doi:10.1016/J.CHROMA.2017.06.013.
580 [31] A. Tuominen, M. Karonen, Variability between organs of proanthocyanidins in *Geranium*
581 *sylvaticum* analyzed by off-line 2-dimensional HPLC-MS, Phytochemistry. 150 (2018) 106–
582 117. doi:10.1016/J.PHYTOCHEM.2018.03.004.
583 [32] J. Hellström, J. Sinkkonen, M. Karonen, P. Mattila, Isolation and Structure Elucidation of
584 Procyanidin Oligomers from Saskatoon Berries (*Amelanchier alnifolia*), J. Agric. Food
585 Chem. 55 (2007) 157–164.
586 [33] I. McMurrough, D. Madigan, M.R. Smyth, Semipreparative Chromatographic Procedure for
587 the Isolation of Dimeric and Trimeric Proanthocyanidins from Barley, J. Agric. Food Chem.
588 44 (1996) 1731–1735. doi:10.1021/jf960139m.

589 [34] M.T. Engström, M. Pälijärvi, C. Fryganas, J.H. Grabber, I. Mueller-Harvey, J.P. Salminen,
590 Rapid qualitative and quantitative analyses of proanthocyanidin oligomers and polymers by
591 UPLC-MS/MS, J. Agric. Food Chem. 62 (2014) 3390–3399. doi:10.1021/jf500745y.
592 [35] J. Suvanto, L. Nohynek, T. Seppänen-Laakso, H. Rischer, J.P. Salminen, R. Puupponen-
593 Pimiä, Variability in the production of tannins and other polyphenols in cell cultures of 12
594 Nordic plant species, Planta. (2017) 1–15. doi:10.1007/s00425-017-2686-8.
595 [36] J.-P. Salminen, M. Karonen, Chemical ecology of tannins and other phenolics: we need a
596 change in approach, Funct. Ecol. 25 (2011) 325–338. doi:10.1111/j.1365-
597 2435.2010.01826.x.
598 [37] J.-P. Salminen, M. Karonen, K. Lempa, J. Liimatainen, J. Sinkkonen, M. Lukkarinen, K.
599 Pihlaja, Characterisation of proanthocyanidin aglycones and glycosides from rose hips by
600 high-performance liquid chromatography-mass spectrometry, and their rapid quantification
601 together with Vitamin C, J. Chromatogr. A. 1077 (2005) 170–180.
602 doi:10.1016/j.chroma.2005.04.073.
603 [38] L. Gu, M.A. Kelm, J.F. Hammerstone, Z. Zhang, G. Beecher, J. Holden, D. Haytowitz, R.L.
604 Prior, Liquid chromatographic/electrospray ionization mass spectrometric studies of
605 proanthocyanidins in foods, J. Mass Spectrom. 38 (2003) 1272–1280. doi:10.1002/jms.541.
606 [39] H.J. Li, M.L. Deinzer, Tandem mass spectrometry for sequencing proanthocyanidins, Anal.
607 Chem. 79 (2007) 1739–1748. doi:10.1021/ac061823v.
608 [40] M. Karonen, J. Liimatainen, J. Sinkkonen, Birch inner bark procyanidins can be resolved
609 with enhanced sensitivity by hydrophilic interaction HPLC-MS, J. Sep. Sci. 34 (2011) 3158–
610 3165. doi:10.1002/jssc.201100569.
611 [41] A.B. Howell, J.D. Reed, C.G. Krueger, R. Winterbottom, D.G. Cunningham, M. Leahy, A-
612 type cranberry proanthocyanidins and uropathogenic bacterial anti-adhesion activity,
613 Phytochemistry. 66 (2005) 2281–2291. doi:10.1016/j.phytochem.2005.05.022.
614

615 **10 Tables**

616 Table 1. The plant species and tissues used in this study.

Plant Species			Sephadex LH-20 fractions 6	
Latin name	Common name	Plant tissue	[a]mDP	[b]PC/PD ratio
Salix phylicifolia L.	Tea-leaved willow	leaves	20	15/85
Pinus sylvestris L.	Scots pine	needles	18	24/76
Aesculus hippocastanum L.	Horse-chestnut	leaves	4	99/1
Larix sp	-	needles	12	62/38
Trifolium repens L.	White clover	flowers	18	2/98
Trifolium medium L.	Zigzag clover	flowers	9	99/1
Ribes alpinum L.	Alpine currant	leaves	19	3/967
Lysimachia vulgaris L.	Garden loosestrife	flowers	20	27/73
Rhododendron schlippenbachii Maxim.	The royal azalea	leaves	6	78/22
Rhododendron dichroanthum Diels.	-	leaves	5	79/21
Lotus corniculatus L.	Birdsfoot trefoil	green brownish pods	14	45/55

[a]Mean degree of polymerization (mDP) and [b]ratio of procyanidin and prodelphinidin subunits (PC/PD) measured by UPLC-MS/MS [32].

617

618 Table 2. Interpretations of the high-resolution mass spectrum of *Lythrum salicaria* semipreparative LC fraction 67. The numbering of the

619 oligomeric series refers to Figure 7.

Oligomer series	monoisotopic *m/z* value of the most abundant oligomer	Charge state	Measured accurate mass (amu)	Mass error (ppm)	PA units in the most abundant oligomer		The range of monomer units per oligomer series	
					PC	PD	PC	PD
DP 2 & 4	584.1233	2	1170.2612	–2.49	3	1	4–0	0–4
	577.1338	1	578.1411	–2.32	2	0	0	2
DP 5	720.1572	2	1442.3289	–2.58	5	0	5–0	0–5
DP 3 & 6	865.1938	1	866.2011	–5.48	3	0	2–0	1–3
	864.1881	2	1730.3907	–3.06	6	0	5–0	1–6
DP 10	1002.5259	3	3010.5995	–3.12	2	8	9–2	1–8
DP 7	1008.2207	2	2018.4559	–1.71	7	0	7–1	0–6
DP 11	1098.5461	3	3298.6602	–3.67	3	8	10–1	1–10
DP 4,8,12	1201.2383	1	1202.2456	–6.97	1	3	4–0	0–4
	1160.2473	2	2322.5092	–3.64	7	1	8–1	0–7
	1205.2291	3	3618.7091	–4.53	2	10	10–0	2–12
DP 13	1301.2484	3	3906.7672	–5.56	3	10	10–2	3–11
DP 9	1352.2616	2	2706.5377	–4.74	2	7	8–1	1–8
DP 14	1402.6007	3	4210.8240	–5.51	3	11	9–1	5–13
DP 5,10,15	1505.2934	1	1506.3007	–7.68	1	4	5–0	0–5
	1496.2919	2	2994.5983	–5.22	3	7	9–2	1–8
	1493.2909	3	4482.8946	–4.70	5	10	11–2	4–13
DP 16	1594.6432	3	4786.9515	–4.69	5	11	8–2	8–14
DP 17	1701.3311	3	5107.0150	–2.38	4	13	9–2	8–14

620

621

622 Table 3. Interpretations of the high–resolution mass spectrum of *Lythrum salicaria* semipreparative LC fraction 91. The numbering of the

623 oligomeric series refers to Figure 8.

Oligomer series	Monoisotopic *m/z* value of the most abundant oligomer	Charge state	Measured accurate mass (amu)	Mass error (ppm)	PA units in the most abundant oligomer		The range of monomer units per oligomer series	
					PC	PD	PC	PD
DP 5	720.1571	2	1442.32879	–2.64	5	0	4–5	1–0
DP 6	864.1881	2	1730.39073	–3.04	6	0	4–6	2–0
DP 10	959.8728	3	2882.64023	–3.23	10	0	9–10	1–0
DP 7	1008.2187	2	2018.45203	–3.64	7	0	5–7	2–0
DP 11	1055.8927	3	3170.69987	–4.12	11	0	9–11	3–0
DP 8 & 12	1152.2482	2	2306.51091	–5.14	8	0	8	0
	1151.9134	3	3458.76197	–4.15	12	0	10–12	2–0
DP 13	1247.9332	3	3746.82143	–4.88	13	0	11–13	2–0
DP 9&18	1296.2794	2	2594.57339	–4.92	9	0	8–9	1–0
	1295.7748	4	5187.12831	–5.47	18	0		
DP 14	1343.9534	3	4034.88200	–5.23	14	0	12–14	2–0
DP 19	1367.7890	4	5475.18507	–6.39	19	0	17–19	2–0
DP 5&10&15	1441.3134	1	1442.32067	–8.27	5	0	5	0
	1440.3091	2	2882.63283	–5.80	10	0	10	0
	1439.9738	3	4322.94311	–5.41	15	0	14–15	1–0
DP 21	1511.8209	4	6051.31251	–5.67	21	0	17–21	4–0
DP 16	1535.9923	3	4610.99864	–6.78	16	0	14–16	2–0
DP 11&22	1584.3389	2	3170.69239	–6.48	11	0	8–11	3–0
	1583.8351	4	6339.36931	–6.46	22	0	18–22	4–0
DP 17	1632.0130	3	4899.06086	–6.62	17	0	15–17	2–0
DP 23	1659.8492	4	6643.42591	–6.42	22	1	19–23	4–0
DP 18	1733.3691	3	5203.12901	–4.34	17	1	17–18	1–0
DP 25	1803.8793	4	7219.54639	–6.78	24	1	24–25	1
DP 19	1824.0519	3	5475.17765	–7.75	19	0	17–19	2–0

DP 20	1920.0731	3	5763.24122	–7.33	20	0	17–20	3–6

624

11 Figures

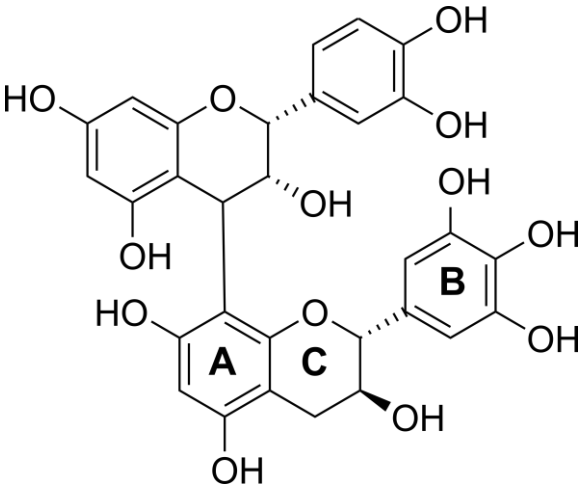
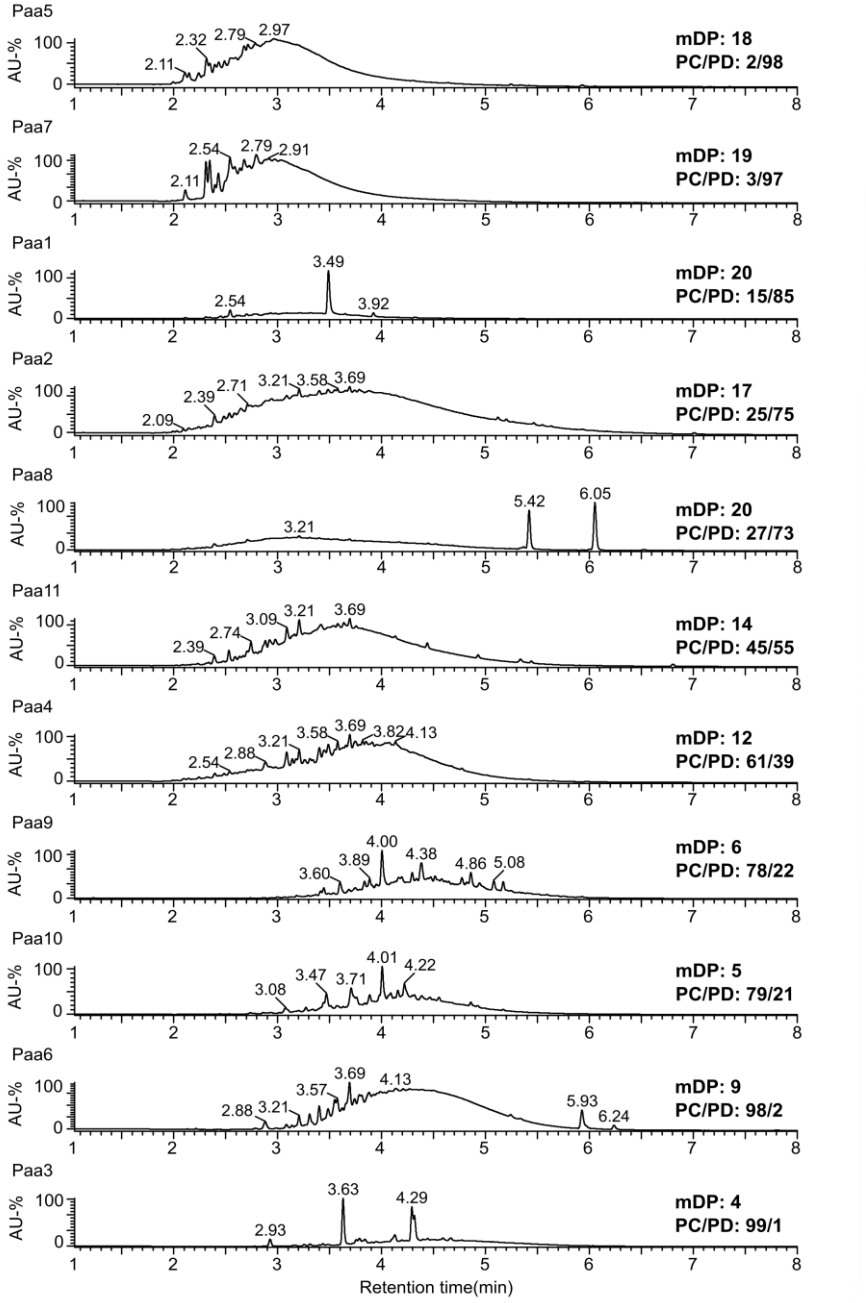


Figure 1. A dimeric proanthocyanidin consisting from the epicatechin extension unit and the gallocatechin terminal unit.



630

631 **Figure 2**. Semipreparative LC traces (λ = 280 nm) of the proanthocyanidin rich 6th Sephadex LH 20

632 fractions.

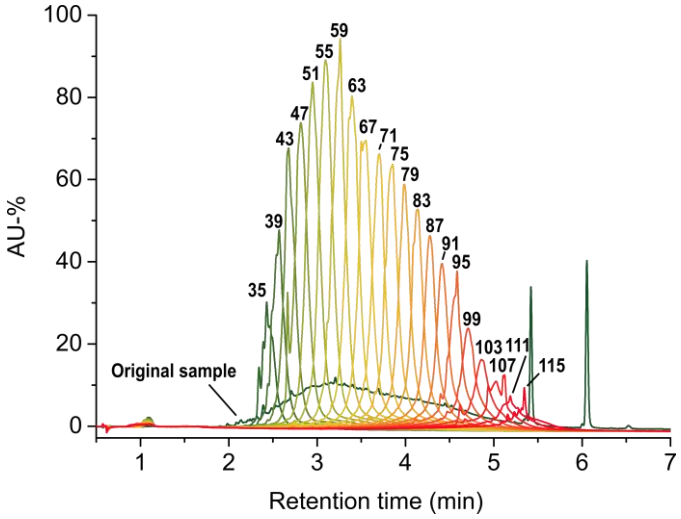
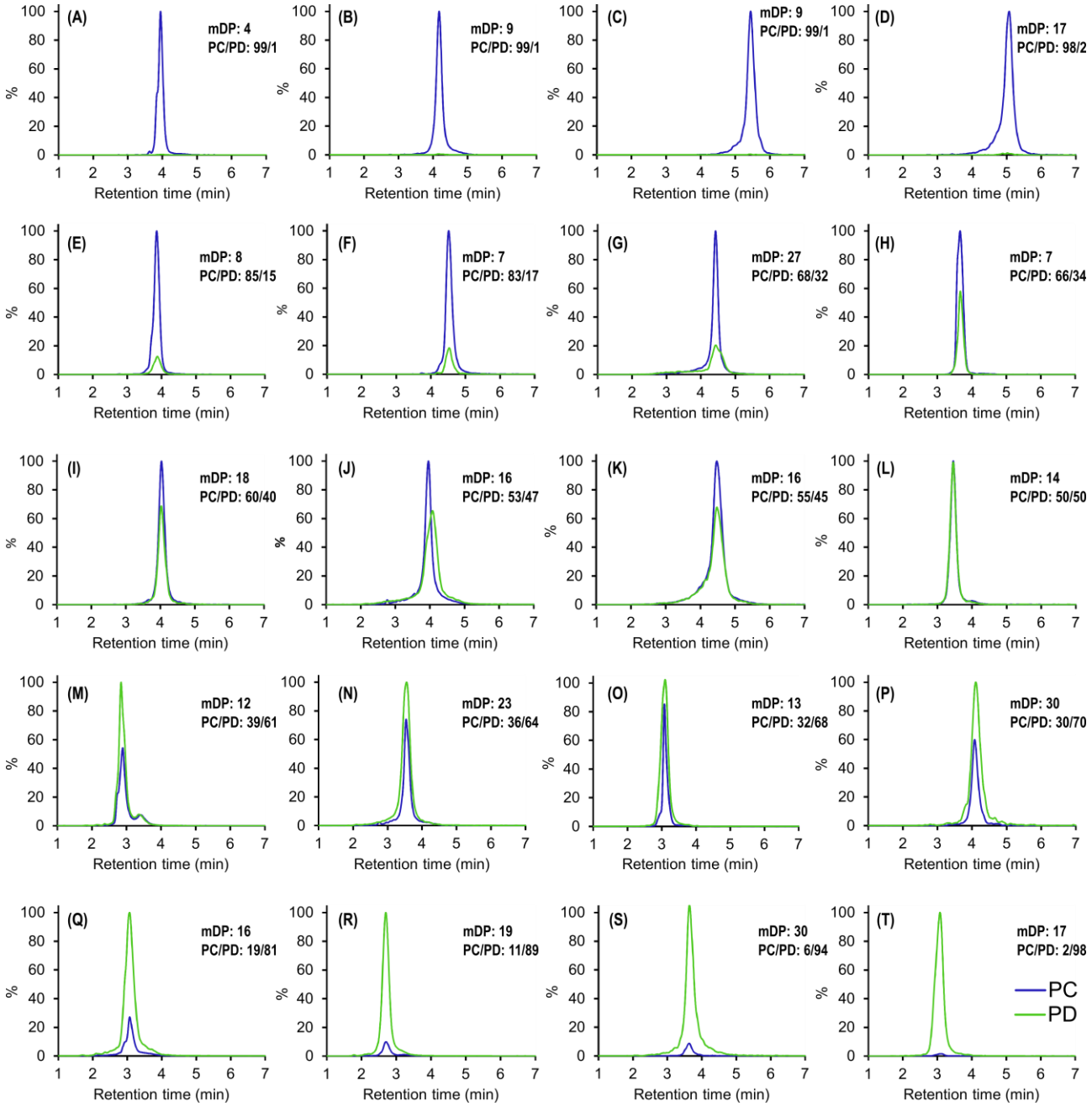


Figure 3. UPLC–DAD chromatogram (λ = 280 nm) of the 6th Sephadex LH-20 fraction of *Lysimachia vulgaris* (original sample) and the further purified semipreparative fractions ranging from fractions 35–115. Color gradient was used to distinguish fractions from each other and the original sample is shown in dark green.



638

639 **Figure 4.** Combined UPLC–DAD–MS/MS single reaction monitoring traces of procyanidin (PC,
640 blue) and prodelphinidin (PD, green) units of semipreparative LC fractions (fr) of (A) *A.*
641 *hippocastanum* fr 76, (B) *T. medium* fr 82, (C) *A. hippocastanum* fr 116, (D) *T. medium* fr
642 106, (E) *Larix sp.* fr 74, (F) *R. dichroanthum* fr 92, (G) *L. vulgaris* fr 91, (H) *R.*
643 *schlippenbachii* fr 70, (I) *L. corniculatus* fr 80, (J) *L. coniculatus* fr 92, (K) *P. sylvestris* fr
644 78, (L) *L.corniculatus* fr 64, (M) *L.corniculatus* fr 48, (N) *L. vulgaris* fr 67, (O) *Larix sp.* fr

54, (P) *S. phylicifolia* fr 80, (Q) *P. sylvestris* fr 54, (R) *L. vulgaris* fr 43, (S) *R. alpinum* fr 68 and (T) *T. repens* fr 54.

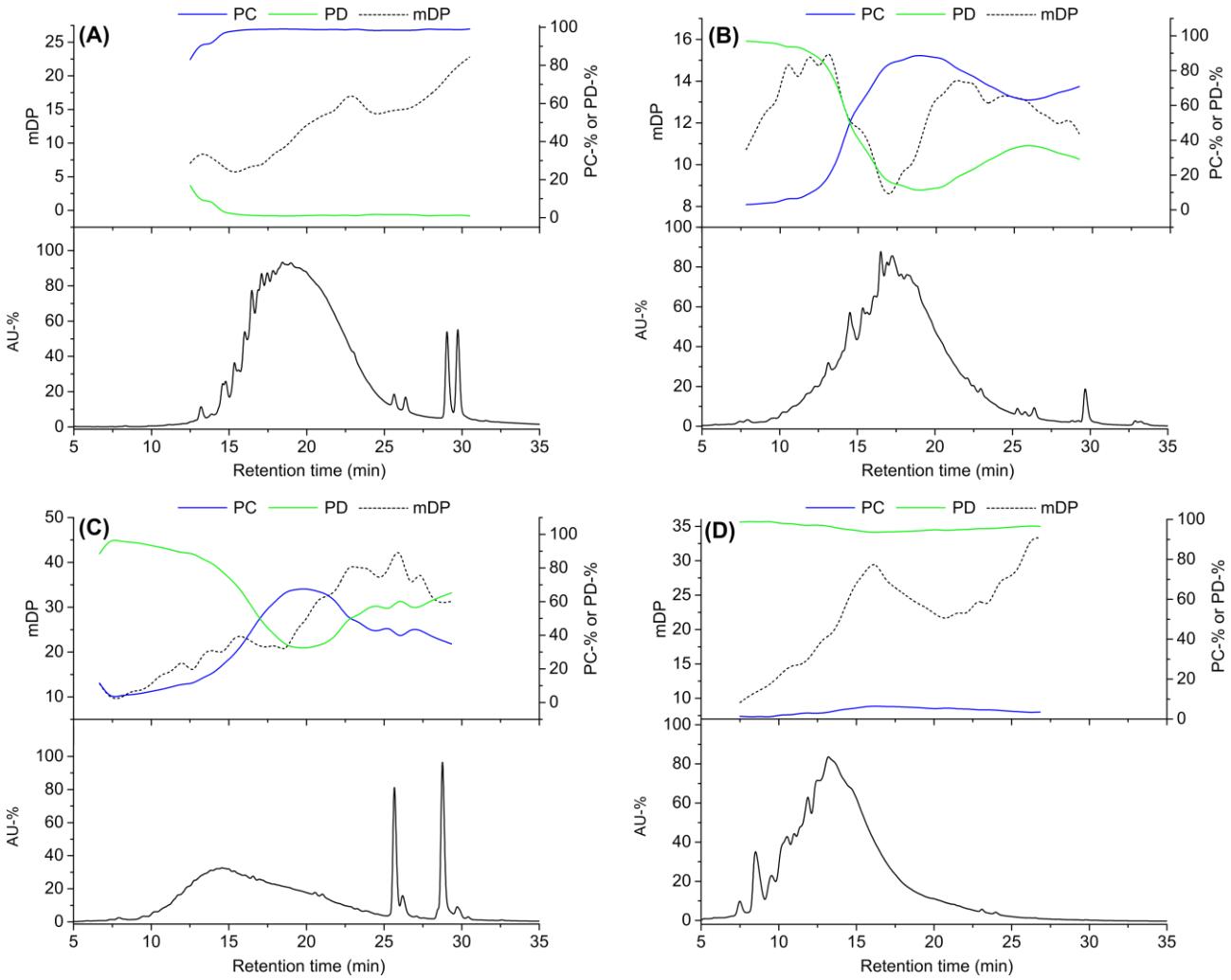
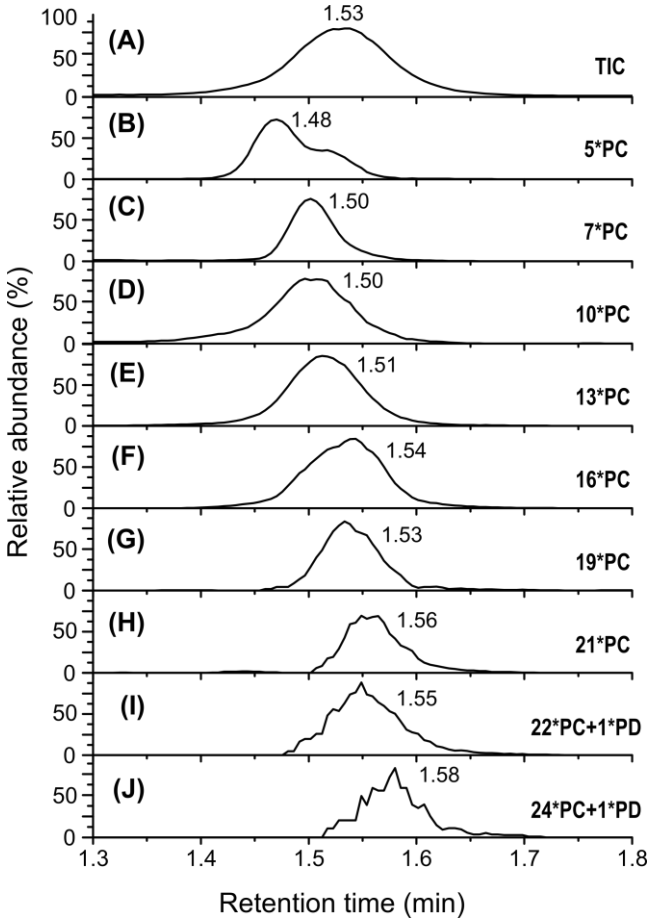


Figure 5. Semipreparative LC UV traces (λ = 280 nm) are displayed in lower panels and the distribution of the mean degree of polymerization (mDP), as well as procyanidin (PC) and prodelphinidin (PD) percentages are displayed in the upper panels. The samples shown are Sep6 fractions of (A) *Trifolium medium*, (B) *Larix sp.*, (C) *Lysimachia vulgaris* and (D) *Ribes alpinum*.



653

654 **Figure 6.** Total ion chromatogram (TIC) and extracted ion chromatograms (EICs) of *Lysimachia*
655 *vulgaris* semipreparative LC fraction 91 obtained by HRMS. The chromatograms are listed
656 as follows: (A) Total ion chromatogram (TIC), and EICs at (B) *m/z* 720.13–720.19, (C) *m/z*
657 1008.18–1008.27, (D) *m/z* 959.83–959.92, (E) *m/z* 1247.88–1247.99, (F) *m/z* 1535.92–
658 1536.08, (G) *m/z* 1823.98–1824.16, (H) *m/z* 1511.75–1511.89, (I) *m/z* 1659.78–1659.92, and
659 (J) *m/z* 1803.80–1803.96. PC refers to procyanidin and PD to prodelphinidin unit, DP refers
660 to degree of polymerization.

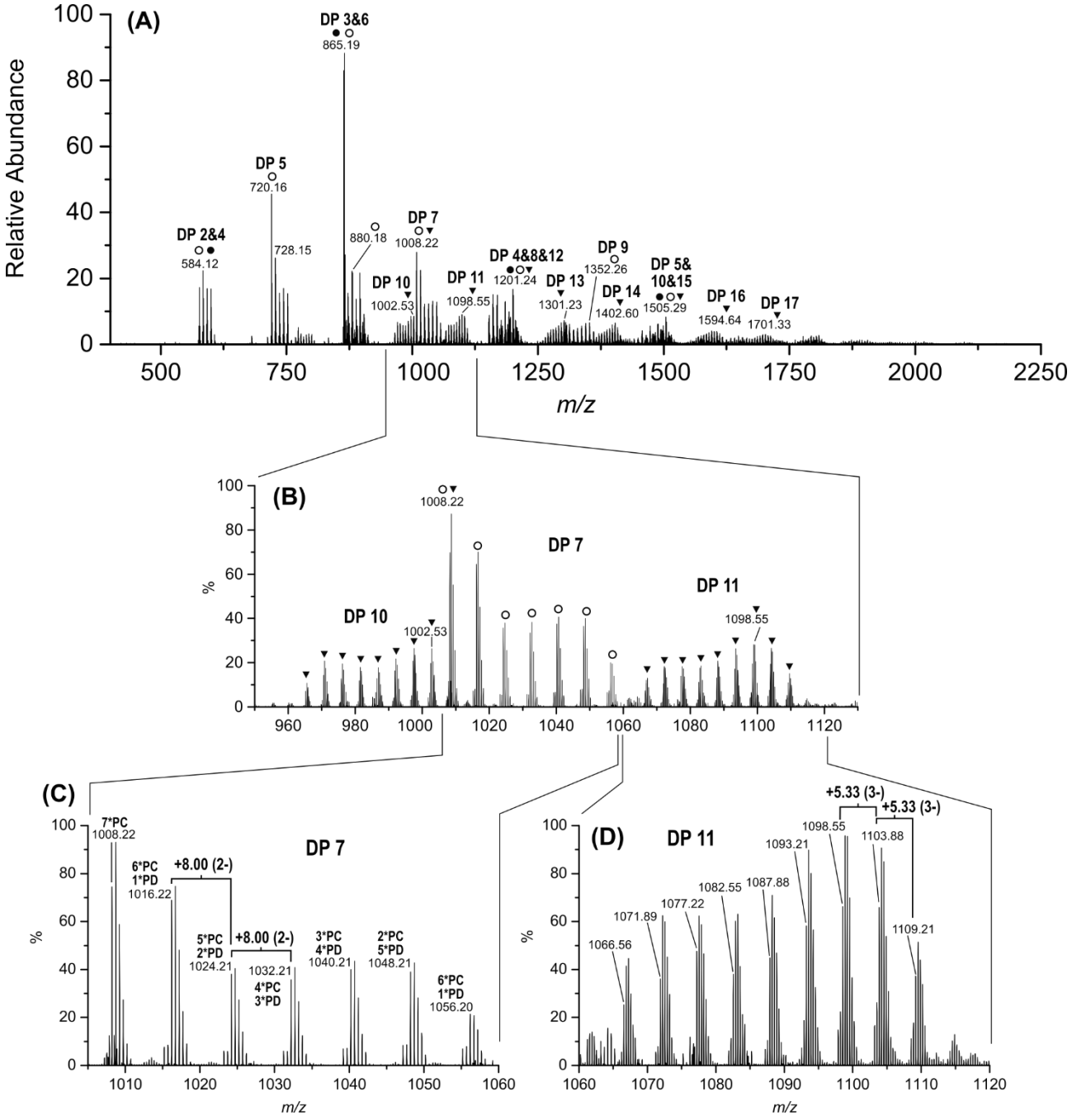


Figure 7. High-resolution mass spectra of *Lysimachia vulgaris* semipreparative LC fraction 67 (A) at *m/z* 400–2250, (B) at *m/z* 950–1125, (C) at *m/z* 1005–1060, (D) at *m/z* 1060–1120. The symbols indicate the charge state of the ions as follows, (●) black dot $[M–H]^{–}$, (○) white dot $[M–2H]^{2–}$, (▼) inverted triangle $[M–3H]^{3–}$. DP refers to degree of polymerization. The interpretations are listed in Table 2.

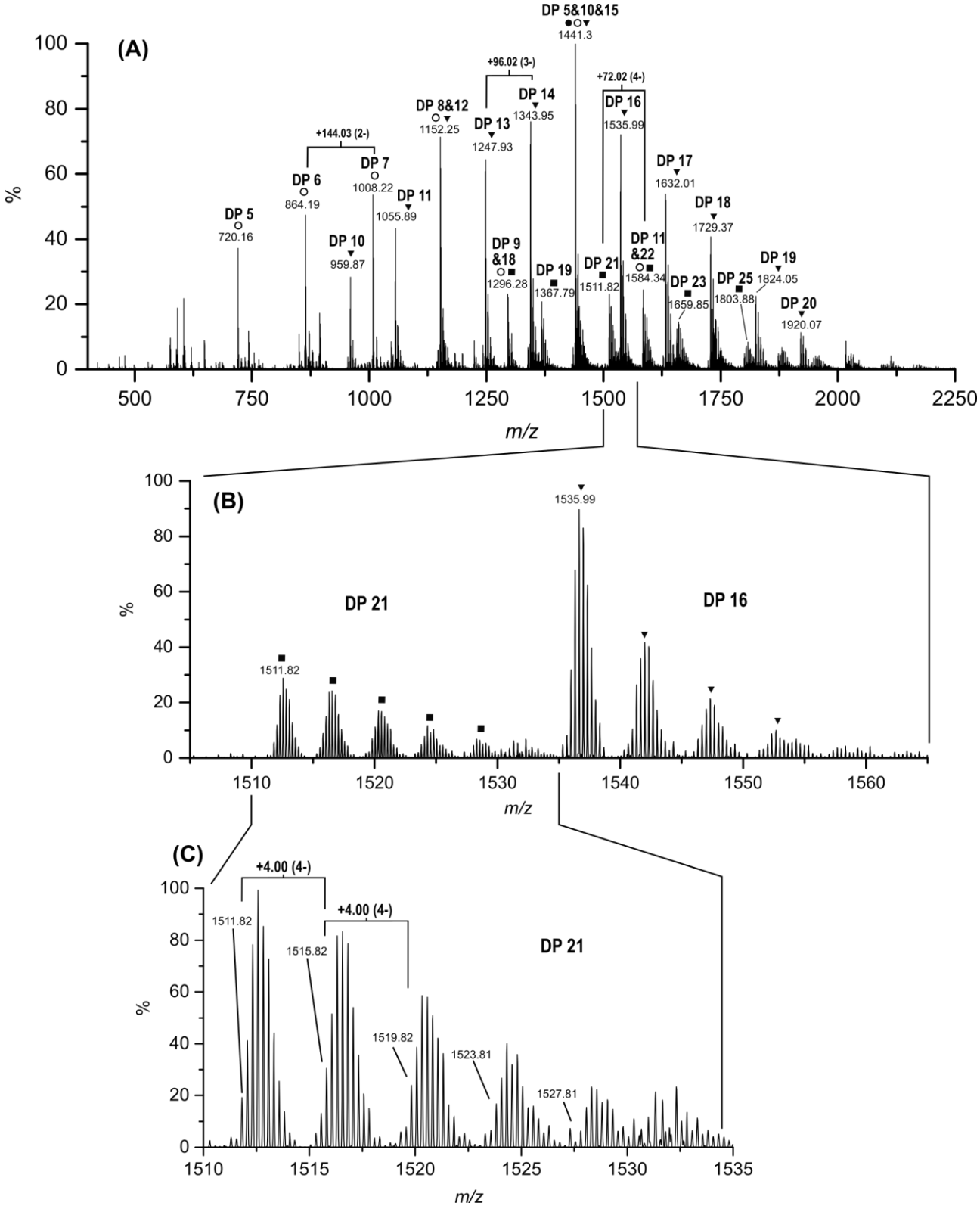
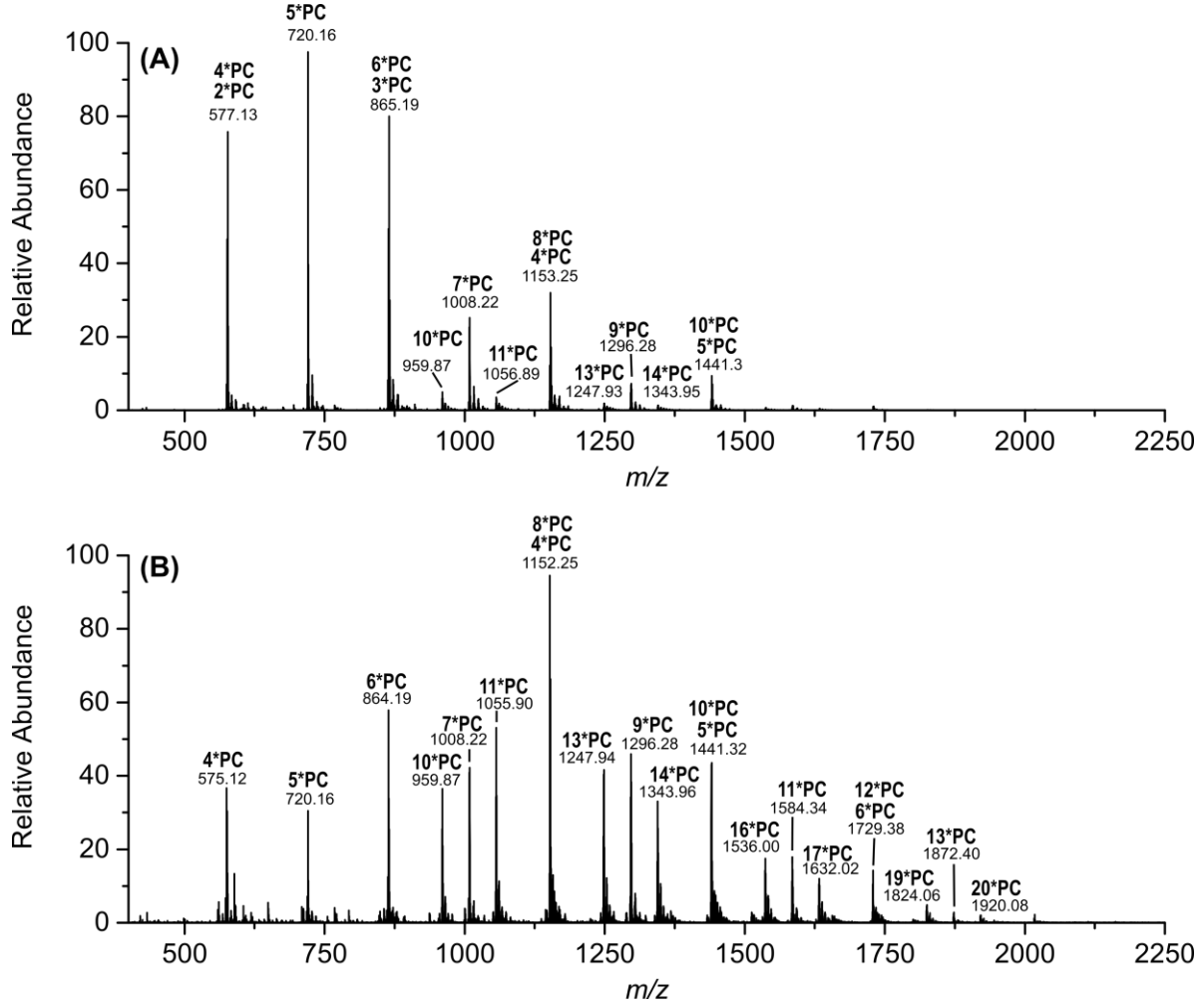


Figure 8. High-resolution mass spectra of *Lysimachia vulgaris* semipreparative LC fraction 91 (A) at *m/z* 400–2250, (B) at *m/z* 1505–1565, (C) at *m/z* 1510–1535 . The symbols indicate the charge state of the ions as follows (○) white dot$[M–2H]^{2–}$, (▼) inverted triangle $[M–3H]^{3–}$,

671 (■) black square $[M-4H]^{4-}$. DP refers to degree of polymerization. The interpretations are

672 listed in Table 3.



673

674 **Figure 9.** HRMS spectra of the *Larix sp.* semipreparative LC fractions (A) 74 and (B) 94.

675

676