Matrix free laser desorption ionization assisted by 13C NMR dereplication: A complementary approach to LC-MS2 based chemometrics

Manon Meunier, Dimitri Bréard, Khalijah Awang, Séverine Boisard, David Guilet, Pascal Richomme, Séverine Derbré, Andreas Schinkovitz

Matrix free laser desorption ionization assisted by ^{13}C NMR dereplication: A complementary approach to LC-MS^2 based chemometrics

Manon Meunier[1], Dimitri Bréard[1,2], Khalijah Awang[3], Séverine Boisard[1,2], David Guilet[1], Pascal Richomme[1], Séverine Derbré[1]*, Andreas Schinkovitz[1]*

[1]University of Angers, SONAS/SFR QUASAV, Campus du végétal 42 rue Georges Morel 49070 Beaucouzé, France

[2]PHYTO platform, SFR QUASAV, Campus du végétal, 49045 Angers CEDEX 01, France

[3]University of Malaya, Faculty of sciences, Department of Chemistry, Malaysia

*corresponding authors

Abstract

The identification of bioactive constituents from complex mixtures of Natural Products (NPs) is an important field of applied chemometrics. It requires the thorough chemical analysis of sample material, which is generally facilitated by liquid chromatography coupled to high-resolution tandem mass spectrometry (LC-MS^2). Providing a simple and time efficient complementary approach, the current work presents the use of matrix free laser desorption ionization mass spectrometry (LDI-MS) within a chemometric model that allows the identification of activity markers in complex mixtures. As a working example the correct prediction and consecutive isolation of Nps with notable anti-AGEs effects from crude bark extracts of *Garcinia parvifolia* is discussed. Results were thoroughly compared to those obtained by a concurrently performed LC-MS^2 analyses. In addition, the ^{13}C NMR dereplication tool MixONat was successfully used to confirm the structures of high yield activity markers directly from complex mixtures prior to their isolation. Overall LDI-MS yielded comparable results to LC-MS^2 and should be considered as complementary technique in applied chemometrics.

Keywords Chemometrics, matrix free laser desorption ionization mass spectrometry, LC-MS^2, ^{13}C NMR dereplication; anti-AGE activity, *Garcinia parvifolia*

1. Introduction

Chemometric analyses have become an important tool in NPs research. These techniques permit the chemical characterization of sample material, but also spotting activity markers within complex mixtures by various statistical means. While *e.g*. Principal Component Analysis (PCA) [1] facilitates the grouping of sample material according to its chemical profile, Partial Least Squares (PLS) regression analysis may further include biological or chemical activity data [2], thus allowing the identification of sample-specific activity markers [3]. Independently from the applied statistical model, all chemometrics do require sound analytical data to yield exploitable results. In this respect, liquid chromatography coupled to high-resolution tandem mass spectrometry (LC-MS^2) is considered a gold standard method. Combining the assets of chromatographic separation with the sensitivity of mass spectrometry, the technique permits the detection of a wide range of NPs at very low concentrations [4]. However, despite these indubitable benefits, LC-MS^2 may still encounter certain limitations. For example, sample solubilization is generally limited to mobile phases, mostly comprising mixtures of acetonitrile, methanol, and water, which are not well adapted to lipophilic compounds. Commonly applied electrospray ionization (ESI) typically promotes the formation of quasi-molecular ions such as $[M+H]^+$ or $[M-H]^-$. However, certain constituents, *e.g.* the lichen quinones haemventosin and parietin, preferably form radical ions and may fail to ionize in ESI [5]. Similar observations were also made for some vitamin E isomers [6]. Facilitating the formation of both, radical and quasimolecular ions [5-7], matrix free laser desorption ionization mass spectrometry (LDI-MS) may therefore provide a valuable supplement to LC-ESI-MS [8]. As many NPs show good absorption at the emission wavelength of common UV-LDI lasers, the method covers a wide range of chemically diverse compounds such as alkaloids [9], phenolics [10, 11] or cannabinoids [12]. Moreover, LDI-MS may be directly applied on crude extracts of different polarities [9, 10] without any sample prepurification steps and even works on powdered non-extracted plant material [13]. In addition, experimental times are much shorter than in LC-MS^2. However, despite these promising results and to the best of the authors' knowledge, LDI-MS has never been included into a chemometric workflow targeting the prediction of (bio)active compounds.

With this in mind, the current manuscript describes the step-by-step development of an LDI-MS based chemometric PLS regression model for the identification of activity markers from complex mixtures of NPs. A fluorometric assay measuring the inhibition of the formation of advanced glycation end-products (AGEs) was used for activity evaluation [14, 15]. In the process, a statistical model based on well-defined mixtures of active and inactive reference compounds was first established (**I**). Next, real-life samples comprising four *Garcinia parvifolia* bark extracts were studied (**II**). Results were then compared to concurrently performed LC-MS^2 based

chemometrics (**III**). Finally, selected activity markers previously identified by LDI-MS and LC-MS^2 were analyzed by MixONat, a free ^{13}C NMR dereplication tool (**IV**) that was developed in the SONAS laboratory [16].

2. Material and methods

2.1. Standard mixtures reference compounds, extraction solvents

Reference compounds amentoflavone (**1**) and betulinic acid (**2**) were purchased from Extrasynthese (Genay, France), while ferulic acid (**3**) was bought from Sigma Aldrich (St. Louis, USA), and 3-hydroxyxanthone (**4**) was obtained from Thermo Fisher (Kandel, Germany). Stock solutions of all compounds were prepared in acetone at a concentration of 10.0 mg/mL. Working solutions were prepared in molar ratios outlined in **Table 1**. A detailed experimental protocol is provided in **Table S1** (supplementary material). Except for acetone (analytical grade), all solvents used for stock solutions, plant extractions, flash chromatography and thin layer chromatography (TLC), were of HPLC grade and purchased from Carlo Erba reagents, (Val-de-Reuil, France).

2.2. Plant collection

Garcinia parvifolia bark samples were collected in Malaysia at four different locations: KL5670, Hutan Simpanan Meranto Gua Musang, Kelantan (date of collection: 13.05.2009). KL5248: Hutan Simpanan Sungai Temau Luala Lipis, Pahang (date of collection: 25.05.2006). KL5259: Km 11 Sanggang, Kuala Kangsar, Perak (date of collection: 22.06.2006). KL5073: 1.2 km Kg. Toh Kah, Jerangau, Terengganu (date of collection: 20.08.2004). Plant material was verified by the botanist Mr. Teo Leong Eng and voucher specimens are kept at the herbarium of the Department of Chemistry, University of Malaya, Kuala Lumpur, Malaysia.

2.3. Extraction procedure

For initial experiments, 10 g of grinded *G. parvifolia* samples were mixed with 100 mL dichloromethane (DCM) and extracted at room temperature for 30 min using an ultrasound bath (Elma, Singen, Germany). The instrument operated at 280 W and at a frequency of 37 kHz. After filtration and solvent evaporation, the following dry weights were obtained: KL5670: 460 mg, KL5248: 470 mg; KL5259: 444 mg, and KL5073: 578 mg. In order to facilitate additional sample processing, KL5670 (80 g) was re-extracted using an up-scaled extraction protocol, which yielded 4.48 g of dry extract.

2.4. Flash chromatography

Sample KL5670 (4.20 g) was fractionated on a CombiFlash Rf flash chromatograph (Serlabo Technologies, Entraigues, France) using a Puriflash 50SIHC-F0220 (220 g) silica gel column. The mobile phase consisted of petroleum ether (solvent A), an equivolumetric mixture of ethyl acetate and chloroform (solvent B), and methanol (solvent C). The flow rate was set to 50 mL/min and the following gradient was applied: 0 min, (A/B/C - 100/0/0); 0-30 min, (A/B/C - 70/30/0); 30-60 min, (A/B/C - 70/30/0); 60-65 min, (A/B/C - 65/35/0); 65-78 min, (A/B/C - 65/35/0); 78-80 min, (A/B/C - 61/39/0); 90-110 min, (A/B/C - 61/39/0); 110-130 min, (A/B/C - 40/60/0); 130-140 min, (A/B/C - 0/100/0); 140-170 min, (A/B/C - 0/100/0); 170-190 min, (A/B/C - 0/90/10); 190-200 min, (A/B/C - 0/80/20); 200-230 min, (A/B/C - 0/80/20). In order to reasonably reduce the number of samples for consecutive LDI-MS and UPLC-MS^2 experiments, fractions were roughly combined according to their analytical TLC profile (data not shown). All TLC experiments were conducted on analytical silica gel 60 TLC plates (Merck, Darmstadt, Germany) comprising a fluorescence indicator at λ: 254 nm. Mobile phases consisted of mixtures of DCM/MeOH and were adapted according to sample requirements. Overall, 14 recombined fractions were obtained (**Figure S1**, supplementary material).

2.5. Experimental preparative HPLC conditions

Fraction 6 (135 mg) and Fraction 10 (270 mg) were dissolved in MeOH (HPLC grade, VWR chemicals, Rosny-sous-Bois, France) before being processed on a Nexera Prep (Shimadzu, Kyoto, Japan) preparative HPLC system using a Pursuit XRs 5 C-18 column (250 x 21.2 mm, 5 µm, Agilent, Santa Clara, California, USA). The mobile phase consisted of HPLC grade water (Solvent A) and acetonitrile (Solvent B, VWR chemicals) and the flow rate was set to 21.2 mL/min. The injection volume was 2 mL.

Chromatograms were recorded at 254 and 280 nm. For F6 the solvent gradient was as follows: 0 min, 45% B; 0-25 min, 50% B; 25-30 min, 70% B; 30-35 min, 100% B; 35-40 min, 100% B. Two repetitive injections (56.4 and 78.6 mg) were performed and yielded 57.0 mg of rubraxanthone (**5**) at a retention time (rt) of 21.0 min.

Fraction 10 was processed using the following solvent gradient: 0 min, 60% B; 0-15min, 65% B; 15-20 min, 70% B; 20-30 min, 100% B; 30-35 min, 100% B. Overall, three repetitive injections (50.4, 94.3 and 120.0 mg) were performed yielding 46.5 mg of isocowanol (**6**, rt: 26.7 min) and 15.4mg of parvixanthone G (**7**, rt: 12.5 min).

2.6. Experimental UPLC-MS conditions

All experiments were conducted at 25°C on an Acquity UPLC system (Waters, Manchester, UK) using a HSS T3 C-18 (2.1 x 100 mm, 1.8 µm) UPLC column (Waters). LC-MS grade water (solvent A) and acetonitrile (solvent B, LC-MS grade, VWR chemicals), both supplemented with 0.1% of formic acid (LC-MS grade, VWR chemicals) were used as mobile phases. Chromatographic conditions for crude extracts and subfractions of KL5670 are outlined below.

Crude extracts: Samples were dissolved in LC-MS grade methanol (VWR chemicals) at a concentration of 0.5 mg/mL and injected in triplicate (injection volume 1 µL). The mobile phase gradient was as follows: 0 min, 30% B; 0-10 min, 80% B; 10-17 min 100% B; 17-20 min, 100% B. The flow rate was 0.3 mL/min.

Subfractions of KL5670. Samples were dissolved in LC-MS grade methanol (VWR chemicals) at a concentration of 0.02 mg/mL and injected in triplicate (injection volume 1 µL). The mobile phase gradient was as follows: 0 min, 40% B; 0-5 min, 80% B; 5-10 min 100% B; 10-12 min, 100% B. The flow was set to 0.4 mL/min.

Mass spectrometry settings for all samples: Mass chromatograms and spectra were recorded on a WatersXevo®G2-S QTOF mass spectrometer (Waters) using negative electrospray ionization. Source parameters were as follows: capillary and cone were set to 0.50 and 40kV, respectively. Desolvation gas flow was 1000 L/h at 36 °C, and cone gas was set to 100 L/h. Source temperature was 120 °C and data were collected in continuum mode at 0.5 scan/s. Instrument calibration was conducted using a leucine-enkephalin solution at a flow rate of 5 µL/min (calibration points: 554.26202 for negative mode). All LC-MS^2 data were acquired in Data Dependent Analysis (DDA) mode. The mass analyzer mass range was set to 50-1200 Da for full scan mode (scan time: 0.5/s). The MS^2 mode was automatically activated once the Total Ion Current (TIC) exceeded 20,000 intensity/s (crude extracts) or 10,000 intensity/s (subfractions of KL5670) but remained in MS mode when TIC was below these thresholds. Collision energy ramps of 20–25 V for low-mass, and 65–85 V for high-mass analytes were employed.

2.7. Experimental LDI-MS conditions

Experiments were conducted on a Spiral-TOF mass spectrometer (JEOL Ltd., Tokyo, Japan), equipped with a pulsed laser operating at 349 nm. Spectra were recorded within a mass range of 50−1200 m/z. Preceding experiments have shown that best results were obtained in negative ionization mode, so all spectra were recorded in high-resolution spiral negative mode. Laser frequency (50-100 Hz), detector gain (49-67 %) and laser power (55-80 %) were individually adapted to samples requirements.

Standard mixtures of reference compounds initially prepared in acetone were dried and redissolved in DCM. Crude *G. parvifolia* extracts as well as subfractions of KL5670 were dissolved in a mixture of DCM/MeOH

(60/40) at a concentration of 10.0 mg/mL. All samples were deposited in quadruplicates (0.5 µL/each) on the sample carrier.

2.8. UPLC-MS^2 data processing and chemometric analysis

Automated data alignment and peak-picking were conducted by the ProgenesisQI software (Waters, Newcastle, UK). Results were thoroughly inspected and manually corrected if necessary (e.g., regrouping of some ions with identic retention times, peak shapes and/or isotopic distribution). An ANOVA tag (p-value ≤ 0.05) was applied to eliminate non-reproducible data. All ions without MS^2 fragmentation were removed and data were then processed by the Progenesis MetaScope dereplication tool. The latter creates theoretical fragmentation patterns from imported structures and compares them with experimental MS^2 data [17]. Putative structural assignments were made using the "Predicted Carbon-13 NMR Data of Natural Products" (PNMRNP) database, which comprises 211,280 NPs [18]. All structures were exported as .sdf, a format that is compatible with ProgenesisQI. Fragmentation tolerance was set to 5 ppm. Based on these parameters ProgenesisQI created a list of possible structures with their respective probability scores.

Finally, a PLS regression analysis based on spectrometric information (m/z values, retention times and peak areas) as well as activity data (AGEs inhibition) was conducted on EZinfo 3.0 for Waters (Umetrics, Umea, Sweden). The Pareto scale setting was applied.

2.9. LDI-MS data processing and chemometric analysis

Spectral data were exported as text files (.txt) using the Tornado MS (JEOL) software. Post-acquisition calibration and automated peak-picking were conducted by the mMass software [19]. Results were thoroughly inspected and manually corrected if necessary. All spectra were normalized by dividing the intensities of individual mass signals by the overall sum of all detected signals within a spectrum. Spectra were then aligned using the MassUp software [20]. Finally, PLS analyses based on LDI-MS (m/z values, normalized signals intensities) and activity data (AGEs inhibition) were performed on EZinfo (Umetrics, Umea, Sweden) using the same parameters as applied for LC-MS^2 data.

2.10. Experimental NMR conditions

Experiments (^{1}H NMR, ^{13}C NMR, DEPT-135, DEPT-90 and 2D NMR) were performed on a JEOL 400 MHz NMR spectrometer (JEOL, Tokyo, Japan) equipped with an inverse 5 mm probe (ROYAL RO5). Chemical shifts (δ_H and δ_C) were expressed as ppm and ^{1}H coupling constants (*J*) are provided in Hz. Depending on their solubility samples were dissolved either in deuterated DMSO-*d*6 or acetone-*d*6 both purchased from Eurisotop, (St-Aubin,

France). For dereplication experiments M1 (10.0 mg) was dissolved in 500 µL DMSO-*d*6, while F6 (50.0 mg) and F10 (40.0 mg) were dissolved in 500 µL acetone-*d*6.

For ^{13}C NMR (100 MHz) spectra, a WALTZ-16 decoupling sequence was used with an acquisition time of 1.04 s (32,768 complex data points) and a relaxation delay of 2 s. The number of acquisitions (512 – 20,000 scans) was individually adapted to sample requirements. A 1 Hz exponential line-broadening filter was applied to each FID prior Fourier transformation. Raw data were processed by MestReNova 12.0.2 (Mestrelab Research, Santiago de Compostela, Spain) and calibrated to solvent peaks at δ_C 39.52 ppm (DMSO-*d*6) or 29.84 ppm (acetone-*d*6) respectively. Manual phasing and baseline correction were conducted. DEPT experiments were aligned with ^{13}C spectra at a given δ_C. Positive ^{13}C NMR and DEPT-90 as well as positive and negative DEPT-135 signals were collected and exported as .csv files.

2.11. ^{13}C-NMR dereplication

All ^{13}C NMR and DEPT data were processed by the MixONat ^{13}C NMR dereplication tool [21, 22]. In preparation of this analysis, reported structures of the *Garcinia* genus were first downloaded from the Dictionary of Natural Products [23], and compiled in a Garcinia database (Garcinia DB). The latter comprised 718 entries, which were saved in the SDF format. Next, a second, xanthones specific database (Xanthone DB) was created from the Universal Natural Products Database (UNPD), which comprises 211,280 NPs [24]. In order to meaningfully reduce the overall number of compounds, UNPD was filtered by a xanthone scaffold using the ACD/NMR Predictors (C and H) and DB software, version 2019 (Toronto, Canada). Finally, 1694 xanthone structures were obtained from UNPD. Their ^{13}C NMR shifts ($\delta_{C\text{-}SDF}$) were then simulated by the ACD/NMR Predictors (C and H) and DB software as well as the CTypeGen routine from MixONat used for differentiating methyl, methylene, methine or quaternary $\delta_{C\text{-}SDF}$ to get DBs compatible with MixONat [25, 26]. Comparing simulated with experimental data (^{13}C NMR, DEPT-135 and 90), MixONat then ranked associated structures with a score from 0 to 1 (where 1 indicates a full match and 0 the absence of any similarity between simulated and experimental spectra). Finally, experimental data of highest ranked candidates were compared with the literature in order to confirm proposed structural assignments.

2.12. Advanced glycation end-products inhibition assay

Experiments were conducted according to a previously described protocol [14, 15]. In short: Stock solutions (SS) of standard mixtures (M1-M6) and *G. parvifolia* bark extracts were prepared in DMSO at a concentration of 10 mg/mL, while single compounds were dissolved at a concentration of 30 mM. These SS were then diluted

with 50 mM phosphate buffer (pH: 7.4) yielding working solutions (WS) at a concentration range of 10^{-2} to 10 mg/mL for standard mixtures and bark extracts. The WS of single compounds were prepared at a concentration range of $3x10^{-2}$ to 30 mM. Next 10 µL of each WS were deposited in 96 black well bottom plates (Fisher Scientific, Illkirch, France) and mixed with 90 µL of a solution containing bovine serum albumine (11 mg/L, Sigma Aldrich, St Quentin Fallavier, France), D-ribose (0.25 M, Acros Organics, Geel, Belgium), and phosphate buffer (50 mM, NaN_3 0.02%, pH 7.4). Plates were then incubated for 24h at 37°C, before being fluorometrically analyzed (λ_{exc}: 335 nm, λ_{em}: 385 nm) on an Infinite M200 plate reader (Tecan, Lyon, France). The formation of AGEs was calculated according to the following formula (FI = fluorescence intensity):

$$\frac{[FI\ (BSA + ribose + sample) - FI\ (BSA + sample)]\ \text{x}\ 100}{FI(BSA + ribose) - FI(BSA)}$$

Aminoguanidine (Sigma Aldrich, St Quentin Fallavier, France) was used as positive, a mixture of BSA + ribose as negative control. A BSA solution served as blank. Results were expressed as IC_{50} values in µg/mL (mixtures and extracts) or µM (single compounds).

All experiments except for KL5259 (insufficient amount of material) were conducted in triplicate. For IC_{50} calculations dose-effect curves were best fitted with a sigmoidal dose-response equation using the Sigma Plot 12.0 software (Systat, Delaware, USA).

3. Results and discussion

3.1. Development of a LDI-MS based chemometric model using defined mixtures of reference compounds

In order to create a working statistical model, six experimental mixtures (M1-M6) (**Fig. S2**, supplementary material) comprising different quantities of four structurally diverse reference compounds were used. All compounds were detectable by LDI-MS (**Fig**. **S3**, supplementary material) and showed very different inhibitory effects on AGEs formation (**Fig**. **1**). Highest activity was observed for the biflavonoid amentoflavone (**1**) (IC_{50}: 23 ± 2 µM), followed by ferulic acid (**3**) (111 ± 12 µM), and 3-hydroxyxanthone (**4**) (124 ± 14 µM), while the IC_{50} of the triterpene betulinic acid (**2**) exceeded 1.0 mM.

Interestingly, M5, which exhibited the best activity of all mixtures, did not contain the highest amount of **1** (M2) (**Table 1**), which was the most active among all single compounds. Moreover, four mixtures M5, M2, M6, and M1 all exhibited rather similar IC_{50} values within a range of 26-38 µg/mL despite their very different relative composition. The situation is quite similar to what is often observed for complex mixtures of NPs such as crude (plant) extracts. These data, however, do not reveal the samples' underlying active principle without further

statistical analysis. For that reason, PLS regression analysis [2] was selected as statistical model to reprocess these data. Unlike PCA, which solely permits the grouping of sample material according to its chemical composition [27], PLS may further include (bio)activity data. The method correlates independent [e.g. spectral data (X)], with dependent variables [activity data, response variable (H)] as well as variables between themselves [28]. Results from this analysis are summarized in **Fig**. **2**.

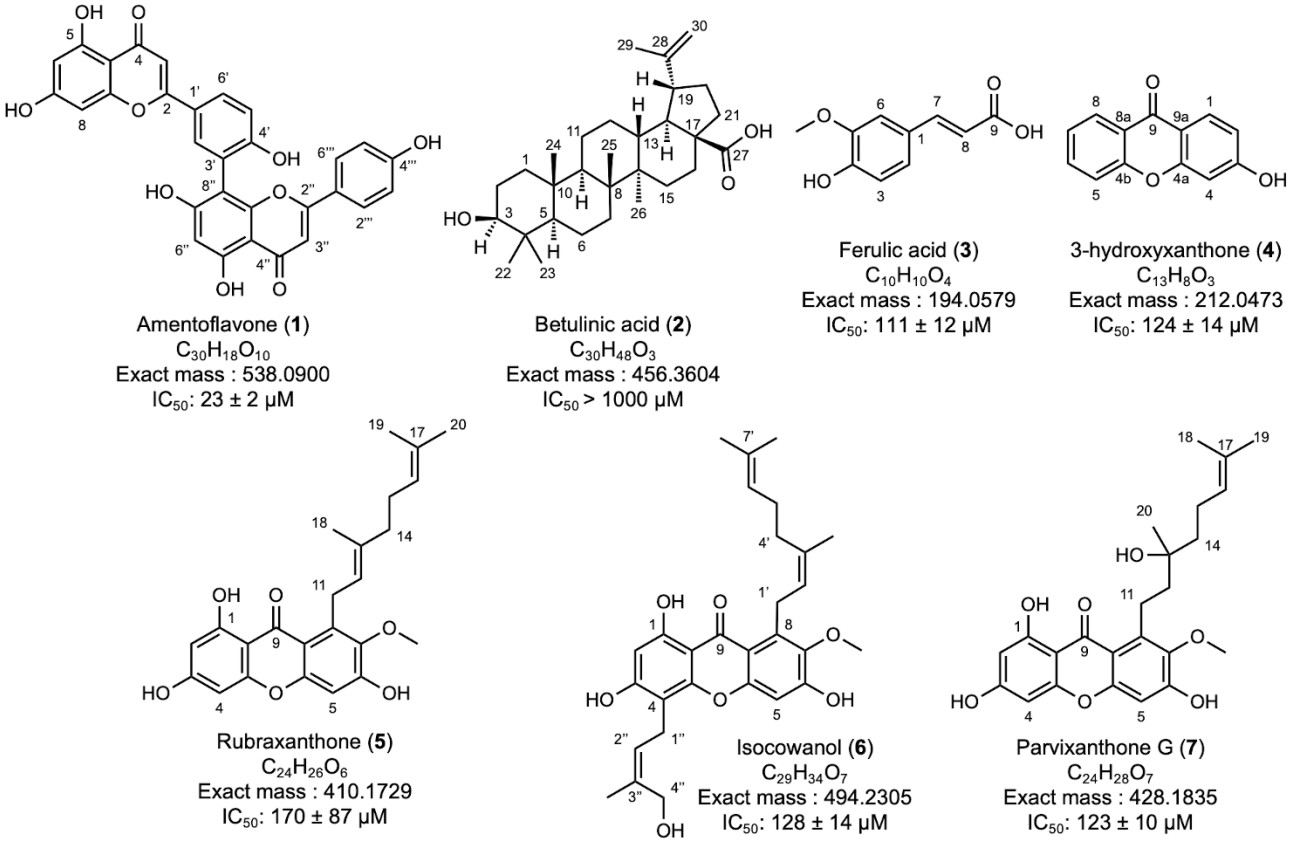


Fig. 1 Structures and anti-AGEs activities of reference (**1**-**4**) and isolated (**5**-**7**) compounds amentoflavone (**1**), betulinic acid (**2**), ferulic acid (**3**), 3-hydroxyxanthone (**4**), rubraxanthone (**5**), isocowanol (**6**) and parvixanthone G (**7**). Averaged IC_{50} values and standard deviations were calculated from three independent experiments. All compounds were detected as quasi-molecular ions $[M-H]^-$ in LDI and ESI-MS. An exhaustive list of theoretical and detected masses is provided in the supplementary material (**Table S2**).

Table 1. Standard mixtures of reference compounds and their anti-AGEs activities. Relative mixing ratios are expressed in mM. Inhibition of AGEs formation is expressed as IC_{50} ± (standard deviation). Aminoguanidine (AG) was used as positive control (IC_{50}: 139 ± 21 µg/mL or 1 900 ± 300 µM).

Standard mixtures	Compounds				AGEs inhibition IC_{50} (µg/mL)
	(1)	(2)	(3)	(4)	
M5	2	3	6	6	26.3 ± 1.5
M2	12	3	6	6	28.3 ± 3.2
M6	7	7	7	7	34.7 ± 3.2
M1	5	5	13	12	38.0 ± 5.0
M4	2	3	32	6	58.0 ± 4.6
M3	2	14	6	6	73.0 ± 5.6
AG					139.0 ± 21.0

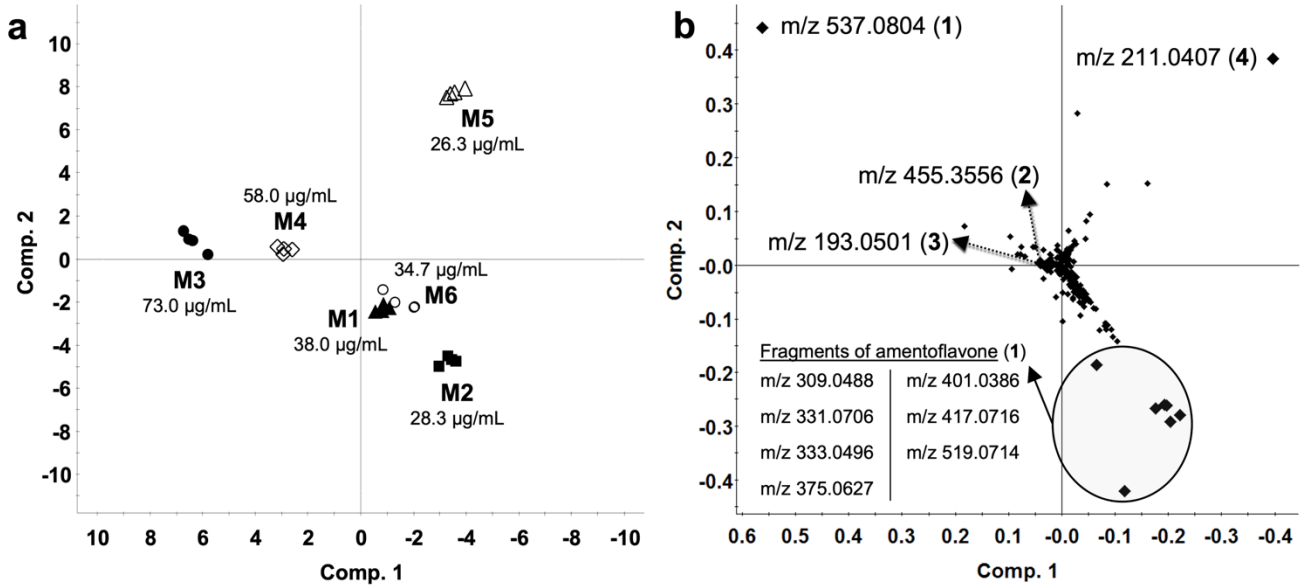


Fig. 2 Partial least square scores plot (**a**) of mixtures M1 to M6 and their anti-AGEs IC_{50} activities. Aminoguanidine was used as positive control (IC_{50}: 139 ± 21 µg/mL). Overall, four principal components accounted for 95.1 % explained variations (1st component: 54.18%; 2nd component: 23.37%; 3rd component: 15.66%; 4th component: 1.89%). The PLS loadings plot (**b**) shows variables of each mixture and molecular ions of reference compounds **1**-**4** are marked as large diamonds and were all detected as $[M-H]^-$. Large diamonds

within the cycle correspond to fragmentation ions of **1**. The rest of the variables (small diamonds) correspond to fragmentation products of compounds **2**-**4**.

Highlighting the good repeatability of LDI-MS, all samples were correctly grouped according to their specific composition (**Fig. 2a**). Samples of highest activities (M5, M2, M6, M1, IC_{50}: 26-38 µM) systematically showed up on the right side of the PLS scores plot, while less active M4 and M3 were found on the left. Most active M5 was shifted to the upper-right corner far off the rest of the samples.

The PLS loadings plot (**Fig. 2b**) then allowed estimating the specific contribution of individual reference compounds to observed anti-AGEs activities of experimental mixtures. According to current results, the farther a marker ion (diamonds) is shifted from the center towards the right side of the graph, the higher its potential contribution to anti-AGEs effects. Moreover, the closer the distance of a marker ion to a specific mixture, the higher its correlation to this sample and its activity. Several marker ions were found in close vicinity of specific mixtures. For example, **4** (211 m/z, IC_{50}: 110.7 µM, upper-right corner) is strongly correlated to M5.

On the other hand, most active **1** (537 m/z, IC_{50}: 22.7 µM) is negatively correlated to M2, which on first sight appears contradictory, as M2 contains the highest quantity of **1** of all mixtures. However, **1** easily fragments in LDI producing fragmentation ions of much higher intensities than the initial non-fragmented compound. These ions are marked by a circle in **Fig. 2b** and correlate very well with M2. Consequently, PLS loadings plots must be always carefully examined with respect to potentially occurring LDI-induced fragmentation.

Compound **3** (193 m/z, IC_{50}: 110.7 µM) is correlated to M4, which is in line with the fact that the compound exhibits highest concentration in specific sample (**Table 1**). For samples M1, M3, M4 and M6, activity correlations to **1**-**4** were generally weak, proposing synergistic effects to explain the observed anti-AGEs effects. Overall, the LDI-MS based PLS loadings plot quickly identified **1** as a major activity marker, despite its strong in-source fragmentation. With regards to unknown samples, the latter may, however, easily lead to misinterpretation of results. Moreover, high resolution LDI-MS provides exact molecular formulas, but it does not allow final structural assignments. Therefore, ^{13}C NMR dereplication may assist the process and provide essential hints for conclusive structure elucidation of major constituents in multi-compound mixtures. In the present case ^{13}C NMR spectra (^{13}C NMR, DEPT-135 and 90) were recorded for M1 (**Fig. S4-S12**, supplementary material) and consecutively processed by the MixONat dereplication software [16]. Based on ^{13}C NMR data of 718 compounds compiled in the Garcinia DB, MixONat correctly predicted the presence of all four reference compounds with the following matching scores: 0.97 for amentoflavone (**1**, rank 3), 0.87 for betulinic acid (**2**, rank 2), 1.00 for ferulic acid (**3**, rank 1), and 0.92 for 3-hydroxyxanthone (**4**, rank 4) (**Fig. S14**, supplementary

material). The dereplication analysis was completed within 30 minutes and experimental data were in line with previous reports (**Table S3**, supplementary material) [29–32].

3.2. The real-life example of *Garcinia parvifolia*

In a next step the combined LDI-MS/^{13}C NMR approach was extended to the chemometric analysis of four bark extracts (KL5670, KL5248, KL5073 and KL5259) of *G. parvifolia*, which all exhibited notable anti-AGEs effects (IC_{50} 79-174 µg/mL, **Fig. 3**). Preceding analyses have further revealed that all extracts comprise significant amounts of xanthones [33]. The latter show close structural similarities to dithranol, an anthraquinone, which is used as a commercial matrix in matrix assisted laser desorption ionization (MALDI) (**Fig. S15**, supplementary material). Therefore, xanthones should easily ionize upon laser irradiation. All extracts were concurrently analyzed by LDI-MS and UPLC-MS^2, and LDI-MS spectra as well as UPLC-MS chromatograms are show in **Fig. S16** and **Fig. S17** respectively (supplementary material). A direct comparison of LDI-MS and UPLC-MS^2 based PLS scores plots presented in **Fig. 3a** & **3c** shows that both methods worked equally well for the differentiation of sample material according to its chemical composition. Overall, three principal components accounted for 98.63% (LDI-MS) and 99.91% (UPLC-MS) of explained variations and 406 (LDI-MS) as well as 7726 (UPLC-MS^2) marker ions were detected. The PLS loadings plot for LDI-MS (**Fig. 3b**) and UPLC-MS^2 (**Fig. 3d**) further revealed that most important activity markers (large diamonds) were commonly detected by the two methods. Noteworthy, while LDI-MS detected one marker ion at 427.1778 m/z (**Fig. 3b**), four ions with practically the same mass (427.1756, 427.1757, 427.1758 and 427.1760 m/z) all accounting for $C_{24}H_{28}O_7$ were observed by UPLC-MS^2 (**Fig. 3d**). This strongly suggests the presence of isomers. The latter can be differentiated by UPLC-MS^2 but not LDI-MS due to the absence of chromatographic separation. Nevertheless, both methods linked activity markers at 409, 427 and 493 m/z to KL5670, as well as 531 m/z to KL5073 (**Fig. 3b** and **3d**).

Considering that KL5670 (IC_{50}: 79.0 ± 7.0 µg/mL) exhibited best anti-AGEs effects of all extracts, the sample was further processed by flash chromatography in order to refine its chemometric analysis. Overall, fourteen subfractions were obtained, evaluated on their anti-AGEs activities, and consecutively analyzed by LDI-MS and UPLC-MS^2. The PLS analyses of these experiments are summarized in **Fig. 4**. As previously observed for crude extracts, LDI-MS and UPLC-MS^2 yielded a very similar grouping of all subfractions in PLS scores plots (**Fig. 4a** & **4c**). Moreover, loadings plots of both methods (**Fig. 4b** & **d**) commonly depicted previously identified activity markers 427 and 493 m/z next to F10 (IC_{50}: 82 ± 18.1 µg/mL), as well as 409 m/z next to F6 (IC_{50}: 58 ± 6.6 µg/mL). In order to obtain further structural information about these compounds, F6 and F10 were analyzed by ^{13}C NMR and MixONat [16, 22, 32]. Next to comparing simulated with reported ^{13}C NMR shifts, the program

further comprises a mass filter, which was set to 410 Dalton when studying F6. MixONat then ranked rubraxanthone (**5**) (hit score 92%) first of the 718 NPs found in the Garcinia DB database. Results were then crosschecked with experimental NMR data from the literature (**Fig. S18**–**S22**, **Table S4**, supplementary material) and further confirmed by ProgenesisQI assisted UPLC-MS^2 data processing. There **5** was ranked in 6th place with a matching score of 55.1% out of 211,000 entries of PNMRNP database (**Table S5**, supplementary material).

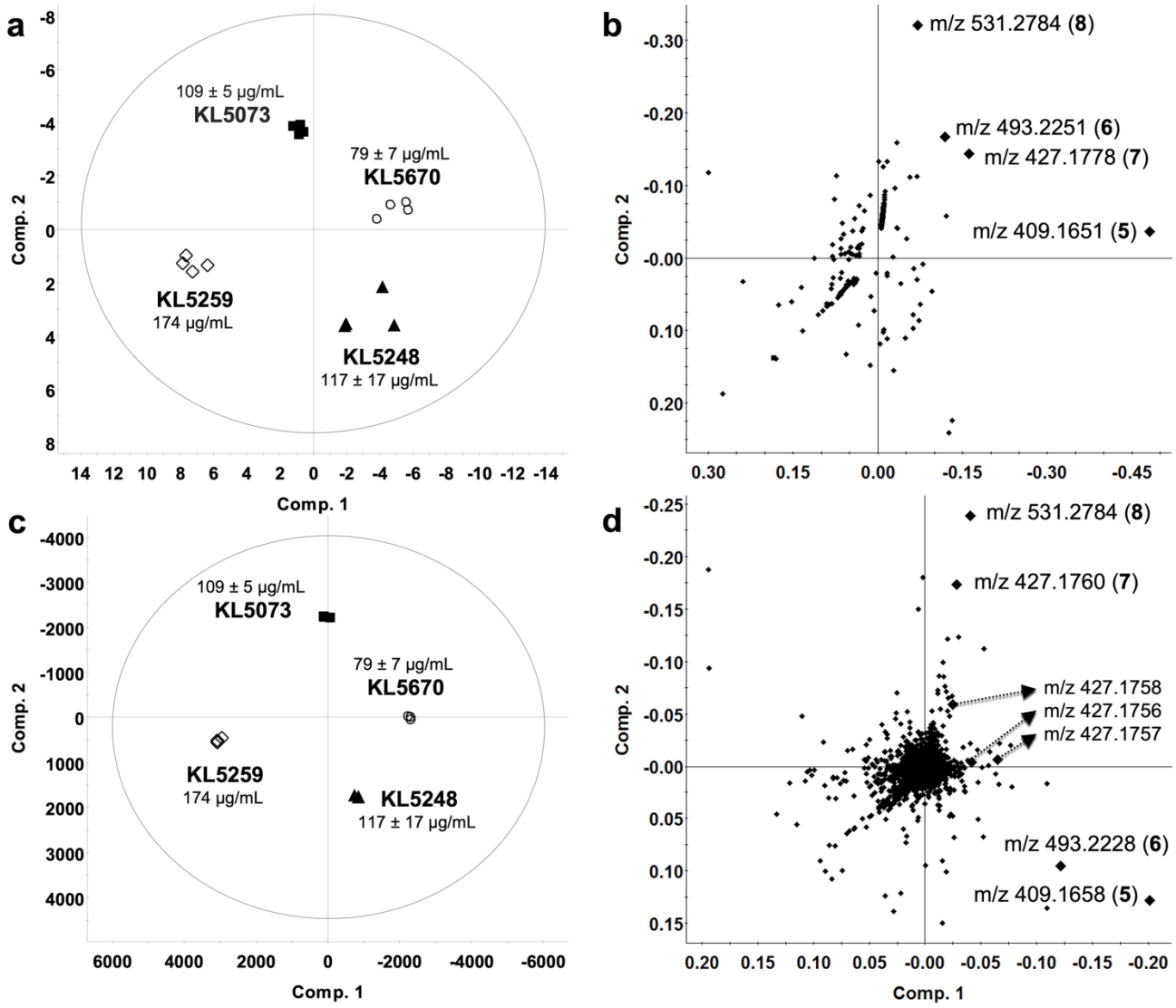
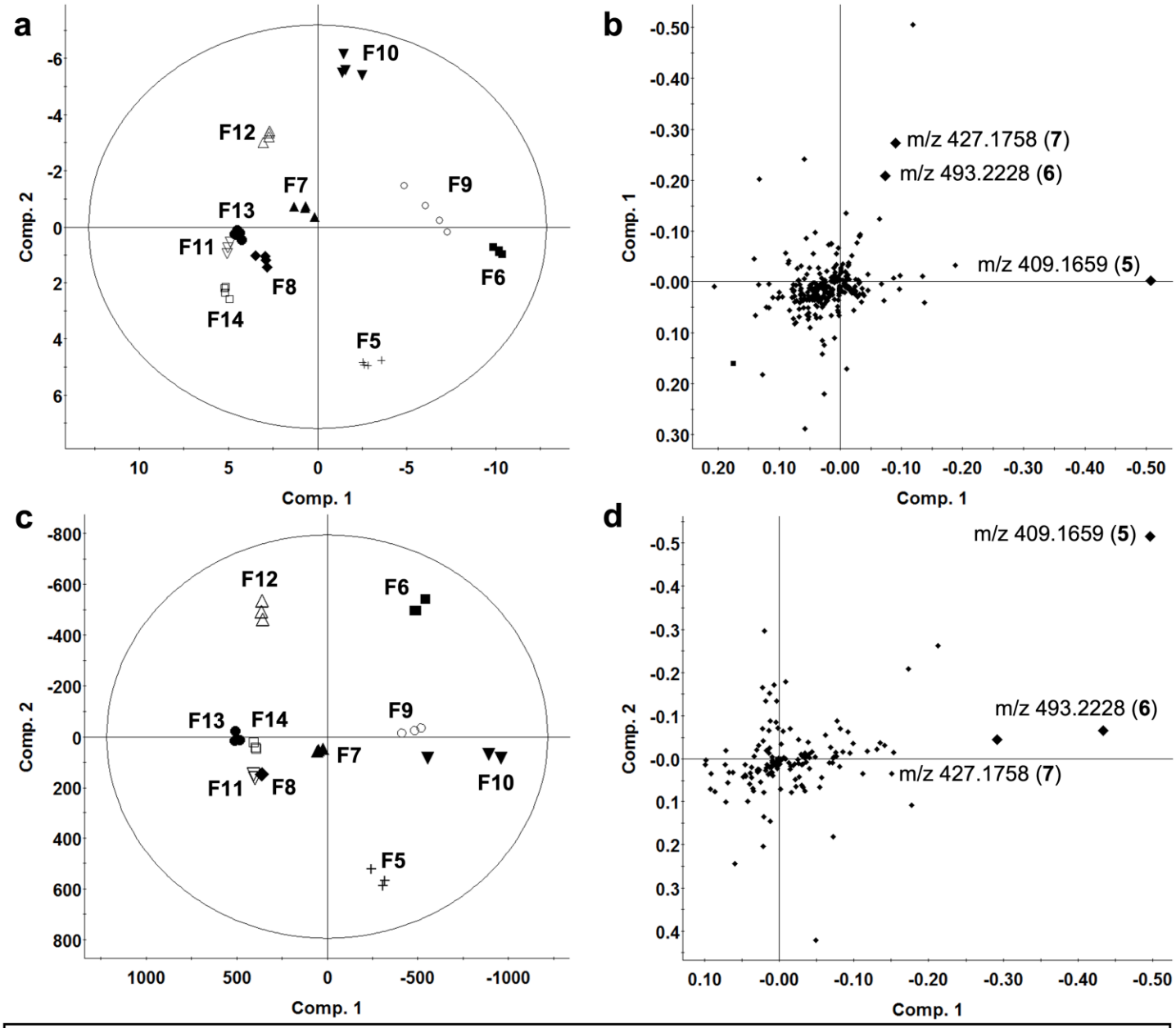


Fig. 3 Partial least square regression analysis of four *G. parvifolia* bark extracts (KL5670, KL5073, KL5259 and KL5248). Both LDI-MS (**a**) and UPLC-MS^2 (**c**) scores plots yielded comparable group separation. Overall, three principal components accounted for 98.63% (LDI-MS) and 99.91% (UPLC-MS^2) of explained variations [1st component: 82.23% (LDI); 94.15% (UPLC-MS^2), 2nd component: 13.88%; 5.64% and 3rd component: 2.52%;

0.12%]. All data showed up within the Hoteling's 95% confidence ellipse. The PLS loadings plots [LDI-MS (**b**) and UPLC-MS^2 (**d**)] proposed identical signals for most prominent activity markers (large diamonds). Based on this information, compounds **5**, **6** and **7** were consecutively isolated and tested (**Fig. 1**).



Anti-AGEs IC_{50} of KL5670 fractions

+ **F5** 145 ± 7 µg/mL	▲ **F7** 104 ± 14 µg/mL	○ **F9** 101 ± 8 µg/mL	▽ **F11** 161 ± 20 µg/mL	● **F13** 147 ± 40 µg/mL
■ **F6** 58 ± 7 µg/mL	◆ **F8** 157 ± 48 µg/mL	▼ **F10** 82 ± 14 µg/mL	△ **F12** 124 ± 27 µg/mL	□ **F14** 161 ± 20 µg/mL

Fig. 4 Partial least square analysis of most active subfraction (F5-F14) obtained after the separation of KL5670 by flash chromatography. As previously observed for crude bark extracts, both LDI-MS (**a**) and UPLC-MS^2 (**c**) yielded very similar sample grouping of all analyzed samples. The PLS loadings plots (**b**, **d**) of both methods mutually linked activity markers isocowanol (**6**) and parvixanthone G (**7**) to F10 and rubraxanthone (**5**) to F6. Overall, three principal components accounted for 98.63% (LDI-MS) and 99.91% (UPLC-MS^2) of explained variations [1st component: 82.23% (LDI); 94.15% (UPLC-MS^2), 2nd component: 13.88%; 5.64% and the 3rd component: 2.52%; 0.12%)]. All data showed up within the Hoteling's 95% confidence ellipse.

For F10 the mass filter was set to 428 and 494 Da and MixONat proposed isocowanol (**6**) and parvixanthone G (**7**) at rank 1 and 3 with hit scores of 86 and 83% respectively (**Fig. S23-S34**, **Table S6**, supplementary material). The ProgenesisQI analysis ranked **6** and **7** in positions 4 and 2 with matching scores of 56.1 and 58.1 respectively (**Table S7-S8**, supplementary material). Finally, both compounds (**6**, 46.5 mg and **7**, 15.4 mg) were isolated by preparative HPLC and analyzed by NMR for final structure confirmation. Results are summarized in the supplementary material (for **6**: **Fig. S35-41**, **Table S7S9**; for **7**: **Fig. S34** and **Table S6-S8**). All isolated marker compounds (**5-7**) exhibited notable anti-AGEs activities in the range 123-170 µM (**Fig. 1**), confirming the proposed LDI-MS/^{13}C NMR based chemometric model.

3.3. ^{13}C NMR assisted LDI-MS complementing UPLC-MS^2 based chemometrics

The example of *G. parvifolia* showed that LDI-MS may yield comparable results to UPLC-MS^2 for the chemical and statistical differentiation of complex mixtures of NPs, provided target analytes ionize well upon laser irradiation. In such a case, LDI-MS may offer several advantages over LC based methods of mass spectrometry: LDI-MS experiments are completed within a few seconds instead of some tens of minutes (LC-MS), hardly require any sample preparations, and are not limited by solvent restriction. On the other hand, in-source fragmentation my complicate the correct interpretation of LDI-MS spectra, a problem that is also known for ESI sources. In absence of chromatographic separation, LDI-MS does not permit the differentiation of isomers, which is exemplified by activity marker ions at 427 m/z in **Fig. 3b** and **3d**. Recent technical advances such as the development of ion mobility mass spectrometry may help bypassing this limitation in the future [34]. At a sufficiently high concentration within sample material, some isomers may further be differentiated ^{13}C NMR.

Compared to LDI-MS, LC-MS^2 generally facilitates the detection of more constituents and compound specific retention times as well as MS^2 data provide additional discriminatory factors for statistical analysis. On the other

hand, both methods may detect different compounds [35] within the same sample, some of which exclusively ionizing in LDI but not in ESI sources [5]. This clearly makes both methods complementary.

Eventually current experiments have shown that LDI-MS data may be easily integrated in multivariate statistics such as PLS, which is a core technique in applied chemometrics. Moreover, predicted activity markers, once isolated, actually showed notable anti-AGEs effects and consequently confirmed the applied statistical model. Next to identifying common activity markers, scores plots (**Fig. 3b** & **3d** and **4b** & **4d**) may further highlight those constituents that are selectively detected by LDI or LC-MS. Consequently, a combined LC-LDI-MS approach may largely extend the range of potential activity markers identified by chemometrics. Provided sufficient material is available material, ^{13}C NMR dereplication tools like MixONat may further assist the process by providing structural evidence for these markers in complex mixtures prior to their isolation. Overall, the presented work highlights the complementarity of LC-MS^2, LDI-MS and ^{13}C NMR as a versatile and holistic approach towards chemometrics in NPs research and other areas of analytical chemistry.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Lever, M. Krzywinski, N. Altman, Principal component analysis, Nat. Methods. 14 (2017) 641–642. https://doi.org/10.1038/nmeth.4346.

[2] D.V. Nica, D.M. Bordean, I. Pet, E. Pet, S. Alda, I. Gergen, A novel exploratory chemometric approach to environmental monitorring by combining block clustering with partial least square (PLS) analysis, Chem. Cent. J. 7 (2013) 145. https://doi.org/10.1186/1752-153X-7-145.

[3] J.J. Kellogg, D.A. Todd, J.M. Egan, H.A. Raja, N.H. Oberlies, O.M. Kvalheim, N.B. Cech, Biochemometrics for natural products research: comparison of data analysis approaches and application to identification of bioactive compounds, J. Nat. Prod. 79 (2016) 376–386. https://doi.org/10.1021/acs.jnatprod.5b01014.

[4] C. Aydoğan, Recent advances and applications in LC-HRMS for food and plant natural products: a critical review, Anal. Bioanal. Chem. 412 (2020) 1973–1991. https://doi.org/10.1007/s00216-019-02328-6.

[5] P. Le Pogam, A. Schinkovitz, B. Legouin, A.C. Le Lamer, J. Boustie, P. Richomme, Matrix-free UV-laser desorption ionization mass spectrometry as a versatile approach for accelerating dereplication studies on lichens, Anal. Chem. 87 (2015) 10421–10428. https://doi.org/10.1021/acs.analchem.5b02531.

[6] L. Zhu, S. Yang, G. Li, X. Zhang, J. Yang, X. Lai, G. Yang, Simultaneous analysis of tocopherols, tocotrienols, phospholipids, γ-oryzanols and β-carotene in rice by ultra-high performance liquid chromatography coupled to a linear ion trap-orbitrap mass spectrometer, Anal. Methods. 8 (2016) 5628–5637. https://doi.org/10.1039/C6AY00556J.

[7] R. Zenobi, R. Knochenmuss, Ion formation in MALDI mass spectrometry, Mass Spectrom. Rev. 17 (1998) 337–366. https://doi.org/10.1002/(SICI)1098-2787(1998)17:5<337::AID-MAS2>3.0.CO;2-S.
[8] F. Chemat, M. Abert-Vian, A.S. Fabiano-Tixier, J. Strube, L. Uhlenbrock, V. Gunjevic, G. Cravotto, Green extraction of natural products. Origins, current status, and future challenges, Trends Anal. Chem. 118 (2019) 248–263. https://doi.org/10.1016/j.trac.2019.05.037.
[9] M. Skopikova, M. Hashimoto, P. Richomme, A. Schinkovitz, Matrix-free laser desorption ionization mass spectrometry as an efficient tool for the rapid detection of opiates in crude extracts of Papaver somniferum, J. Agric. Food Chem. 68 (2020) 884–891. https://doi.org/10.1021/acs.jafc.9b05153.
[10] P. Le Pogam, P. Richomme, M.A. Beniddir, T.H. Duong, G. Bernadat, A. Schinkovitz, A thorough evaluation of matrix-free laser desorption ionization on structurally diverse alkaloids and their direct detection in plant extracts, Anal. Bioanal. Chem. 412 (2020) 7405–7416. https://doi.org/10.1007/s00216-020-02872-6.
[11] A. Schinkovitz, S. Boisard, I. Freuze, J. Osuga, N. Mehlmer, T. Brück, P. Richomme, Matrix-free laser desorption ionization mass spectrometry as a functional tool for the analysis and differentiation of complex phenolic mixtures in propolis: a new approach to quality control, Anal. Bioanal. Chem. 410 (2018) 6187–6195. https://doi.org/10.1007/s00216-018-1225-1.
[12] N.A. dos Santos, L.M. de Souza, F.E. Pinto, C. de J. Macrino, C.M. de Almeida, B.B. Merlo, P.R. Filgueiras, R.S. Ortiz, R. Mohana-Borges, W. Romão, LDI and MALDI-FT-ICR imaging MS in Cannabis leaves: optimization and study of spatial distribution of cannabinoids, Anal. Methods. 11 (2019) 1757–1764. https://doi.org/10.1039/C9AY00226J.
[13] S. Islam, R. Alam, S. Kim, Improved coverage of plant metabolites using powder laser desorption/ionization coupled with Fourier-transform ion cyclotron mass spectrometry, Food Chem. 373 (2022) 131541. https://doi.org/10.1016/j.foodchem.2021.131541.
[14] S. Derbré, J. Gatto, A. Pelleray, L. Coulon, D. Séraphin, P. Richomme, Automating a 96-well microtiter plate assay for identification of AGEs inhibitors or inducers: application to the screening of a small natural compounds library, Anal. Bioanal. Chem. 398 (2010) 1747–1758. https://doi.org/10.1007/s00216-010-4065-1.
[15] L. Séro, L. Sanguinet, P. Blanchard, B. Dang, S. Morel, P. Richomme, D. Séraphin, S. Derbré, Tuning a 96-well microtiter plate fluorescence-based assay to identify AGE inhibitors in crude plant extracts, Molecules. 18 (2013) 14320–14339. https://doi.org/10.3390/molecules181114320.
[16] A. Bruguière, S. Derbré, J. Dietsch, J. Leguy, V. Rahier, Q. Pottier, S. Suor-Cherer, G. Viault, MixONat, a software for the dereplication of mixture based on ^{13}C NMR spectroscopy, Anal. Chem. 92 (2020) 8793–8801. https://doi.org/10.1021/acs.analchem.0c00193.
[17] X. Shi, W. Yang, S. Qiu, J. Hou, W. Wu, D. Guo, Systematic profiling and comparison of the lipidomes from Panax ginseng , P. quinquefolius , and P. notoginseng by ultrahigh performance supercritical fluid chromatography/high-resolution mass spectrometry and ion mobility-derived collision cross section measurement, J. Chromatogr. A. 1548 (2018) 64–75. https://doi.org/10.1016/j.chroma.2018.03.025.
[18] M. Lianza, R. Leroy, C. Machado Rodrigues, N. Borie, C. Sayagh, S. Remy, S. Kuhn, J.H. Renault, J.M. Nuzillard, The three pillars of natural product dereplication. Alkaloids from the bulbs of Urceolina peruviana (C. Presl) J.F. Macbr. as a preliminary test case, Molecules. 26 (2021) 637. https://doi.org/10.3390/molecules26030637.
[19] M. Strohalm, M. Hassman, B. Košata, M. Kodíček, mMass data miner: an open source alternative for mass spectrometric data analysis, Rapid Commun. Mass Spectrom. 22 (2008) 905–908. https://doi.org/10.1002/rcm.3444.
[20] H. López-Fernández, H.M. Santos, J.L. Capelo, F. Fdez-Riverola, D. Glez-Peña, M. Reboiro-Jato, Mass-up: an all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery, BMC Bioinformatics. 16 (2015) 318. https://doi.org/10.1186/s12859-015-0752-4.

[21] A. Bruguière, S. Derbré, D. Bréard, F. Tomi, J.M. Nuzillard, P. Richomme, ^{13}C NMR dereplication using MixONat software: a practical guide to decipher natural products mixtures, Planta Med. 87 (2021) 1061–1068. https://doi.org/10.1055/a-1470-0446.
[22] A. Bruguière, S. Derbré, P. Richomme, MixONat, SourceForge. (2020). https://sourceforge.net/projects/mixonat/ (accessed January 14, 2022).
[23] Dictionary of natural products 26.2 chemical search, ChemNetBase. (2015). https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml (accessed November 1, 2015).
[24] P. M. Allard, T. Péresse, J. Bisson, K. Gindro, L. Marcourt, V.C. Pham, F. Roussi, M. Litaudon, J.L. Wolfender, Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication, Anal. Chem. 88 (2016) 3317–3323. https://doi.org/10.1021/acs.analchem.5b04804.
[25] A. Bruguière, S. Derbré, C. Coste, M. Le Bot, B. Siegler, S.T. Leong, S.N. Sulaiman, K. Awang, P. Richomme, ^{13}C-NMR dereplication of Garcinia extracts: predicted chemical shifts as reliable databases, Fitoterapia. 131 (2018) 59–64. https://doi.org/10.1016/j.fitote.2018.10.003.
[26] L.F. Silva-Castro, S. Derbré, A.M. Le Ray, P. Richomme, K. García-Sosa, L.M. Peña-Rodriguez, Using ^{13}C-NMR dereplication to aid in the identification of xanthones present in the stem bark extract of Calophyllum brasiliense, Phytochem. Anal. 32 (2021) 1102–1109. https://doi.org/10.1002/pca.3051.
[27] E.R. Britton, J.J. Kellogg, O.M. Kvalheim, N.B. Cech, Biochemometrics to identify synergists and additives from botanical medicines: a case study with Hydrastis canadensis (Goldenseal), J. Nat. Prod. 81 (2018) 484–493. https://doi.org/10.1021/acs.jnatprod.7b00654.
[28] L. Ory, E.-H. Nazih, S. Daoud, J. Mocquard, M. Bourjot, L. Margueritte, M.A. Delsuc, J.M. Bard, Y.F. Pouchus, S. Bertrand, C. Roullier, Targeting bioactive compounds in natural extracts - Development of a comprehensive workflow combining chemical and biological data, Anal. Chim. Acta. 1070 (2019) 29–42. https://doi.org/10.1016/j.aca.2019.04.038.
[29] Z. Qunfang, X. Zhanhui, T. Pengfei, L. Gansun, C. Hongming, A new triterpene from rosemary (Rosmarinus officinalis), J. Chin. Pharm. Sci. 9 (2000) 131–133.
[30] M. Xiang, H. Su, J. Hu, Y. Yan, Isolation, identification and determination of methyl caffeate, ethyl caffeate and other phenolic compounds from Polygonum amplexicaule var. sinense, J. Med. Plants Res. 5 (2011) 1685–1691.
[31] S.H. Goh, I. Jantan, P.G. Waterman, Neoflavonoid and biflavonoid constituents of Calophyllum inophylloide, J. Nat. Prod. 55 (1992) 1415–1420. https://doi.org/10.1021/np50088a005.
[32] E.G.R. Fernandes, A.M.S. Silva, A.S. Cavaleiro, F.M. Silva, M.F.M. Borges, M.M. Pinto, ^{1}H and ^{13}C NMR spectroscopy of mono-, di-, tri- and tetrasubstituted xanthones, Magn. Reson. Chem. 36 (1998) 305–309. https://doi.org/10.1002/(SICI)1097-458X(199804)36:4<305::AID-OMR193>3.0.CO;2-N
[33] Y.J. Xu, Y.H. Lai, Z. Imiyabir, S.H. Goh, Xanthones from Garcinia parvifolia, J. Nat. Prod. 64 (2001) 1191–1195. https://doi.org/10.1021/np0101393.
[34] K. Masike, M.A. Stander, A. de Villiers, Recent applications of ion mobility spectrometry in natural product research, J. Pharm. Biomed. Anal. 195 (2021) 113846. https://doi.org/10.1016/j.jpba.2020.113846.
[35] G. Petroselli, M.K. Mandal, L.C. Chen, G.T. Ruiz, E. Wolcan, K. Hiraoka, H. Nonami, R. Erra-Balsells, Mass spectrometry of rhenium complexes: a comparative study by using LDI-MS, MALDI-MS, PESI-MS and ESI-MS: LDI, MALDI, PESI and ESI-MS, J. Mass. Spectrom. 47 (2012) 313–321. https://doi.org/10.1002/jms.2965.