Title: A hexa-herbal TCM decoction used to treat skin inflammation: An LC-MS-based phytochemical analysis

Jennifer B. Chang[1], Dr. Majella E. Lane[2], Dr. Min Yang[3], Prof. Michael Heinrich[1]

Affiliation

[1] Research Cluster Biodiversity and Medicines /Centre for Pharmacognosy and Phytotherapy, UCL School of Pharmacy, London, U.K.

[2]Department of Pharmaceutics, UCL School of Pharmacy, London, U.K.

[3]Department of Pharmaceutical & Biological Chemistry, UCL School of Pharmacy, London

Corresponding author:

Prof. Michael Heinrich, Research Cluster Biodiversity and Medicines/Centre for Pharmacognosy and Phytotherapy, UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX, U.K. E-mail: m.heinrich@ucl.ac.uk Phone: +44 (0) 207 753 5844

*Dedicated to Professor Dr. Dr. h.c. mult. Kurt Hostettmann in recognition of his outstanding contribution to natural product research

Abstract

In order to understand the chemical relationship between a traditional hexa-herbal Chinese medicine formula (HHCF) and botanical drugs it is derived from, an analytical platform comprising of liquid chromatography coupled with triple quadrupole mass spectrometry (LC-MS/MS) and data mining was developed to separate and identify key chemical components. The HHCF comprises the rootstock of *Scutellaria baicalensis* Georgi (SCU), *Rheum tanguticum* Maxim. ex Balf. (RHE), *Sophora flavescens* Aiton (SOP); root bark of *Dictamnus dasycarpus* Turcz. (DIC); bark of *Phellodendron chinense* C.K. Schneid. (PHE) and fruit *of Kochia scoparia* (L.) Schrad. (KOC). 73 compounds including alkaloids, anthraquinone derivatives, coumarins, coumarins derivatives, flavonoids, flavone glycosides, naphthalene derivatives, phenylbutanone glucopyranoside, phenolic acids, pterocarpans, stilbenes, stilbenes derivatives and tannins were putatively identified based on mass measurement and characteristic fragment ions. Among the botanical drugs of the HHCF, RHE, SOP, PHE and SCU contributed the majority of the extracted metabolites of the HHCF decoction. The developed method appeared to be a versatile tool for monitoring chemical constituents in extracts of a TCM formula in a relatively comprehensive and systematic manner and help to understand the importance of the individual botanical drugs within a formulation.

Key words

data mining, *Dictamnus dasycarpus* Turcz., *Kochia scoparia* (L.) Schrad., LC-MS/MS, *Phellodendron chinense* C.K. Schneid., phytochemical profiling, *Rheum tanguticum* Maxim. ex Balf., *Scutellaria baicalensis* Georgi, *Sophora flavescens* Aiton, TCM formula,

Abbreviations

BDD	Botanical drug decoction
DIC	Root bark of *Dictamnus dasycarpus* Turcz.
E_t	End time
HHCF	Hexa-herbal Chinese formula
KOC	Fruit *of Kochia scoparia* (L.) Schrad.(KOC)
PHE	Bark of *Phellodendron chinense* C.K. Schneid.
RHE	Rootstock of *Rheum tanguticum* Maxim. ex Balf.
R_t	Retention time
SCU	Rootstock of *Scutellaria baicalensis* Georgi
SOP	Rootstock of *Sophora flavescens* Aiton
S_t	Start time
Traditional Chinese medicine	TCM

Introduction

Chinese herbal medicines are generally prescribed as a formula comprised of multiple botanical drugs that are expected to exert clinical effects based on the combined effects of multiple components against multiple targets. [1] Characterization of the chemical components in a Chinese herbal preparation is thus vital for understanding its pharmacological mechanism and for achieving reliable clinical effects. Liquid chromatography combined with triple quadrupole mass spectrometry (LC-MS/MS) is a versatile tool for profiling chemical components in Chinese herbal formulations with its separation power and mass measurement. [2-4]

To understand the chemical relationship between a traditional Chinese medicine (TCM) formula and the botanical drugs it is derived from, LC-MS/MS was used in combination with an in-house developed template to identify chemical components extracted from each botanical drug in a formula. The example is a hexa-herbal Chinese formula (HHCF) comprising of rootstock of *Scutellaria baicalensis* Georgi (SCU), *Rheum tanguticum* Maxim. Ex Balf. (RHE), *Sophora flavescens* Aiton (SOP); root's bark of *Dictamnus dasycarpus* Turcz. (DIC); bark of *Phellodendron chinense* C.K. Schneid. (PHE) and fruit *of Kochia scoparia* (L.) Schrad. (KOC). This HHCF is formulated to treat skin inflammation associated with pathogenic factors i.e. wind, dampness and heat.

Results and Discussion

Method development and Data mining

The total ion chromatograms (TIC) of the HHCF (Fig. S1), DIC (Fig. S2), KOC (Fig. S3), PHE (Fig. S4), RHE (Fig. S5), SCU (Fig. S6) and SOP (Fig. S7) were obtained by LC-MS/MS and analyzed using an in-house developed template, concepts employed in the template were illustrated in Fig. 1. The in-house developed template defined sources and possible identities of ions detected in the HHCF decoction by carrying out cross-matching of retention time-mass pair of ions detected in the HHCF and botanical drug decoctions (BDD), followed by matching the mass of detected ions against the mass of reported compounds from their source.

Identification of compounds

Three cases were encountered during identification of compounds detected in the HHCF decoction, 1) mass of a detected ion matched with mass of one reported compound; 2) mass of a detected ion matched with the mass of multiple reported compounds and 3) multiple detected ions having same mass (different elution time) matched against mass of multiple reported compounds. For case 1 and 2, the identification of compounds was putatively assigned based on the observed fragmentation patterns illustrated in Fig. 2 (for the top five most abundant compounds in positive and negative ionization mode), S8 (all compounds putatively identified in positive ion mode) and S9 (all compounds putatively identified in negative ion mode). Detected ions in case 3 having the same mass and similar fragment ions; their identification was putatively ascribed based on observed fragment patterns and reported retention behavior of these compounds in reversed-phase chromatography. Compounds were putatively identified in this study and falls into level 2 or 3 of the four levels identification rigor defined by the Metabolomics Standards Initiative. [5] Overall, evidences considered

during the assignment of compound identification involved mass of precursor ion, elution behaviour, relatively intensities of fragment ions (Table S2 and S3) or characteristic MS/MS fragments that were compare against reported MS/MS spectra (Table S2 and S3).

Based on the mass measurement of precursor and fragment ions, 33 and 40 compounds in the HHCF decoction were putatively identified in positive (See Table 1 and Fig. S8) and negative ionization mode (See Table 2 and Fig. S9), respectively.

Compound P6, P7, P10, P18 and P23 gave identical $[M+H]^+$ ions at *m/z* 265 in the TIC and further yielded similar fragment ions in the MS/MS analysis. They were considered to be isomers extracted from SOP with a molecular weight of 264 e.g. 5α-hydroxymatrine (Fig.S8, P6/P10), 9α-hydroxymatrine (Fig. S8, P6/10), 14β-hydroxymatrine (Fig.S8, P7), oxymatrine (Fig. S8, P18/P23) and oxysophoridine (Fig.S8, P18/P23) or lamprolobine (Fig.S10). [6-8] The observed fragment ions at *m/z* 247 for all these compounds excluded the possibility of them being lamprolobine. Compound P18 and P23 gave and additional $[2M+H]^+$ ions at *m/z* 529 in the TIC and yielded a fragment ion at *m/z* 205 that characterizes them to be oxy-alkaloids i.e. oxymatrine and oxysophoridine. [6,7] Fragment ions at *m/z* 127 were only observed in Compound P7 that characterized the cleavage of bonds of 5-17 and 7-11 and the presence of a hydroxyl group at position 14 of 14β-hydroxymatrine. Compound P6 and P10 were putatively identified as 5α-hydroxymatrine or 9α-hydroxymatrine and the observed fragment ions could not distinguished the position of the hydroxyl group.

Compound P11, P15 and P20 gave identical $[M+H]^+$ ions at *m/z* 247 in the TIC and further yielded similar fragment ions in the MS/MS analysis. They were considered to be isomers extracted from SOP with a molecular weight of 246 e.g. 5, 6-dehydrolupanine (Fig.S8, P11),

isosophocarpine (Fig.S8, P15), sophocarpine (Fig.S8, P15), or 7,11-dehydromatrine (Fig. S8, P20). [6,7] Compound P15 yielded prominent fragment ions at *m/z* 179 and other ions at *m/z* 150, 148, 136 and 227 that were reported to be fragment ions observed for isosophocarpine or sophocarpine. [6] Thus, compound P15 could be either isosophocarpine or sophocarpine. Fragment ions at *m/z* 179 were observed in the MS/MS analysis of compound P20 but not in compound P11, which characterized cleavage of bonds of 11-12 and 15-16 of 7,11-dehydromatrin. In addition, fragment ions at *m/z* 179 are unlikely to be observed in the MS/MS analysis of 5,6-dehydrolupanine. Compound P20 and P11 was therefore putatively assigned to be 7, 11-dehydromatrine and 5, 6-dehydrolupanine, respectively.

Compound P12, P17, P21 and P22 gave identical $[M+H]^+$ ions at *m/z* 263 in the TIC and further yielded similar fragment ions in the MS/MS analysis. They were considered to be isomers extracted from SOP with a molecular weight of 262 e.g. mamanine (Fig.S8, P12), oxysophocarpine (Fig.S8, P17), 9α-hydroxysophocarpine (Fig.S8, P21), leontalbinine N-oxide (Fig.S8, P22) and (-)-9α-hydroxy-7, 11-dehydromatrine (Fig. S11). [6,8] Compound P17 also gave $[2M+H]^+$ ions at *m/z* 525 in the TIC that characterized it to be oxy-alkaloids i.e. oxysophocarpine. [6,7] In addition, fragment ions at *m/z* 195 were seen in the MS/MS analysis of compounds P17 and P21, which characterized cleavage of bonds of 11-12 and 15-16 and the presence of a double bond between position 13 and 14 of oxysophocarpine or 9α-hydroxysophocarpine. Compound P17 and P21 were therefore putatively identified as oxysophocarpine and 9α-hydroxysophocarpine, respectively. Similar to compounds P17, P22 gave additional $[2M+H]^+$ ions at *m/z* 525 in the TIC that characterized it to be oxy-alkaloids i.e. leontalbinine N-oxide. The MS/MS analysis of compound P12 produced fragment ions at *m/z* 128 that suggested it to be mamanine.

Compound P14 and P16 gave identical $[M+H]^+$ ions at *m/z* 249, $[2M+H]^+$ ions at *m/z* 497 and $[2M+Na]^+$ ions at *m/z* 519 in the TIC. Compound P19 gave $[M+NH_4]^+$ ions at *m/z* 266 in the TIC. Compound P14, P16 and P19 further yielded similar fragment ions that were considered to be isomers extracted from SOP with a molecular weight of 248 e.g. allomatrine (Fig. S8, P14/16), isomatrine (Fig. S8, P14/16), matrine (Fig. S8, P14/P16), sophoridine (Fig. S8, P14/P16) or lupanine (Fig. S8, P19). [6-9] The exhibition of both $[M+H]^+$ ions $[2M+Na]^+$ ions provided further support that compound P14 and P16 may be alkaloids with no substituting groups. [7] MS/MS analysis of compound P19 gave fragment ions at *m/z* 205 that was not observed for compound P14 and P16. With reference to reported fragment ions and retention behavior of these compounds in reversed-phase chromatography, [6] Compound P14 and P16 were putatively assigned to be stereoiosmers i.e. allomatrine, isomatrine, matrine or sophoridine and compound P19 was putatively ascribed as lupanine.

Compound P24 sourced from DIC produced no fragment ion in the MS/MS analysis under the tested conditions. Compound P24 with precursor ions at *m/z* 243 could be dasycarpusenester A or O-ethylnor-γ-fagarine (Fig. S9, N24) that exhibited as adduct ions $[M+H]^+$ and $[M]^+$ in the TIC, respectively. [10,11]

Compound N2 and N4 gave identical $[M-H]^-$ ions at *m/z* 191 in the TIC. Compound N2 yielded prominent fragment ions at *m/z* 191, 127 and other ions such as *m/z* 173, 111 and 109 due to elimination of neutral molecules of CO, CO_2 and/or H_2O from *m/z* 191. Compound N4 yielded prominent fragment ions at *m/z* 111 and other ions at *m/z* 155 , 131 and 129 due to losses of neutral molecules of CO_2, OH and/or H_2O from *m/z* 191. Based on literature data, quinic acid yielded addition and base ion at *m/z* 127, while citric acid yielded base ion at *m/z* 111. [14] In addition, it was reported that quinic acid was eluted before citric acid in

reversed-phase chromatography. [15] Therefore, compound N2 and N4 were putatively ascribed to be quinic acid (Fig.S9, N2) and citric acid (Fig.S9, N4), respectively.

Compounds N15 and N21 gave identical $[M-H]^-$ ions at *m/z* 367 in the TIC and were sourced from PHE. The MS/MS fragmentation behavior of these ions was characterized by the elimination of the [quinic acid-H]$^-$ ion at *m/z* 191, [quinic acid-H_2O-H]$^-$ ion at *m/z* 173 and [ferulic acid-H]$^-$ ion at 193. [16] According to previous reports, the linkage position of acyl groups on the quinic acid of these positional isomers were determined based on the base ion produced in the MS/MS analysis. [17,18] Compound N15 and N21 were thus putatively ascribed as 3-*O*-feruloyl quinic acid (Fig.S9, N15) and 5-*O*-feruloyl quinic acid (Fig.S9, N21) on the basis of base ions at *m/z* 193 and 191, respectively.

Compound N27 and N30 gave identical $[M-H]^-$ ions at *m/z* 477 in the TIC and further yielded similar fragment ions. They were indicated to be isomers extracted from RHE with a molecular weight of 478 e.g. isolindleyin (Fig.S9, N27) and lindleyin (Fig.S9, N30). The fragment ions at *m/z* 313 and 169 suggested that compound N27 and N30 were composed of a rheosmin, a glucose and a galloyl units. After examining the retention behavior of these compounds in reversed-phase chromatography, compound N27 and N30 were putatively determined to be isolindleyin and lindleyin, respectively. [3,13]

Compound N29 and N32 gave identical $[M-H]^-$ ions at *m/z* 547 in the TIC and further yielded similar fragment ions in the MS/MS analysis. They were considered to be isomers extracted from SOP with a molecular weight of 548 e.g. chrysin 6-C-arabinosyl 8-C-glucoside (Fig.S9, N29) or chrysin-6-C-glucosyl-8-C-arabinoside (Fig.S9, N32). Both compounds yielded

fragment ions at *m/z* 427 that indicated the presence of C-glycoside unit attached to the flavone skeleton. Fragment ions at *m/z* and 337 suggested the presence of two sugar units. After examining reported retention behavior of these compounds in reversed-phase chromatography, compound N29 and compound N32 were putatively identified as chrysin 6-C-arabinosyl 8-C-glucoside and chrysin-6-C-glucosyl-8-C-arabinoside, respectively. [19,20]

Compound N33 and N34 gave identical $[M-H]^-$ ions at *m/z* 541 in the TIC and further yielded similar fragment ions. They were indicated to be isomers extracted from RHE with a molecular weight of 542 e.g. resveratrol-4'-O-β-D-(2"-O-galloyl)-glucoside (Fig. S9, N33) or resveratrol-4'-O-β-D-(6"-O-galloyl)-glucoside (Fig. S9, N34). Fragment ions at *m/z* 313 and 169 characterized neutral loss of a resveratrol unit *(m/z*=227) and the eliminated galloyl unit, respectively. Based in additional on the reported retention behavior of these compounds in reversed-phase chromatography, compound N33 and N34 were putatively identified as resveratrol-4'-O-β-D-(2"-O-galloyl)-glucoside and resveratrol-4'-O-β-D-(6"-O-galloyl)-glucoside, respectively. [12,13]

n conclusion, PHE, RHE and SOP contributed the largest number of secondary metabolites identified in the HHCF, while KOC seems to contribute very little. The developed analytical platform allowed us to simultaneously analyze multiple chemical constituents in a TCM decoction in a relative comprehensive and systematic manner. The chemical constituents in the TCM decoction could be monitored by means of mass, retention time and relatively abundance of ions in the TIC. This strategies enables the identification of compounds which are likely to be relatively in complex (multi-herbal) preparations.

Materials and methods

Chemicals and materials

All medicinal herbs were purchased from commercial Chinese herbal medicine stores in China. SCU, RHE, SOP, DIC, KOC and PHE were sourced from Hebei (Chengde), Gansu (Maqu county), Hebei (Chengde), Liaoning (Anshan), Hebei (Chengde) and Sichuan (Dujiangyan), respectively, and were authenticated by the first author based on her experience with Chinese herbal medicines. Samples are deposited at the Centre's Herbarium and are numbered as JC1-6. MS grade formic acid and LC-MS grade acetonitrile were obtained from Sigma-Aldrich and LC-MS grade water was obtained from Fisher Scientific.

Preparation of HHCF, SCU, RHE, SOP, DIC, PHE and KOC decoctions for LC-MS/MS analysis

All botanical drugs, except SCU were blended into powder and SCU were cut into small blocks of 1 cm x 1 cm before the decoction process. For the HHCF decoction, the same ratio of each botanical drug (i.e. SCU, RHE, SOP, DIC, PHE and KOC) was used. Medicinal herbs were first macerated in distilled water (at a volume of 5 folds the dry weight of medicinal herbs used) for 1 hour and then heated under reflux for 95 minutes. The extracted solution was filtered through a nylon cloth of pore size ~0.1 mm, followed by centrifugation at 10,000 rpm for 5 minutes. Collected supernatant was lyophilized. Lyophilized extracts were dissolved in LC-MS graded water to achieve a concentration of 20 mg/mL (w/v), centrifuged at 10 000 rpm for 10 minutes and filtered through 0.22 μm filter membrane before analysis.

LC-MS/MS analysis

Chromatography was performed on a 1260 Infinity liquid chromatography system. 15 μL of

sample is injected on to a Zorbax SB-C18 column (4.6 x 250mm, 5 μm) held at 35 °C. The elution program was as follows: 0 min, 5% B; 110 min, 47% B; 120-125 min, 90% B; 126-135 min, 5% B. The flow rate was 1 mL/min, where A was 0.1% formic acid in water and B was 0.1% formic acid in acetonitrile. The LC system was coupled to an Agilent 6400 series Triple Quadrupole mass spectrometer equipped with an Agilent Jet Stream electrospray ionization source (AJS ESI), operating with 3.5 kV capillary voltage in both positive and negative modes (sheath gas temperature 250 °C, sheath gas flow rate 11 L/min, drying gas temperature 300°C, drying gas flow rate 5 L/min and nebulizer pressure 45 psi in both positive and negative mode). The fragmentor voltage is 135V. Positive and negative ionization mode mass spectra were simultaneously collected across the mass range of 100-1500 *m/z*. For MS/MS acquisition mode, the parameters were the same except that collision energy settings of 15V was used.

Data mining

Total ion chromatograms (both positive and negative ion mode) were obtained for the HHCF and its botanical drugs, peaks were extracted and integrated in the Agilent MassHunter Qualitative Analysis software. Start time (S_t), end time (E_t) and retention time (R_t) of mass (*m/z*) and abundance of ions within each peaks (ions within the same peak with abundance < 1% of the abundance of the base ion were excluded) were extracted from the total ion chromatogram.

To define the source of ions extracted in the HHCF decoctions and to match the mass (*m/z*) of ions detected in the HHCF against the database of their source, an in-house developed

template was created by inputting functions into an excel spread sheets to run the data mining steps with concepts illustrated in Fig. 1.

Supporting information

Relative intensity of fragment ions produced in the MS/MS analysis for Table 1 and 2 are available in Table S2 and S3, respectively. Fragmentation pathways of compounds putatively identified in Table 1 and 2 are illustrated in Fig. S8 and S9, respectively. TIC of HHCF, DIC, KOC, PHE, RHE, SCU, SOP are available in Fig. S1-7, respectively. Ions present in the HHCF and more than one composing botanical drugs are listed in Table S1. Chemical structure of lamprolobine and (-)-9α-hydroxy-7,11-dehydromatrine are illustrated in Fig. S10 and S11, respectively.

Acknowledgements

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] *Che CT, Wang ZJ, Chow MSS, Lam CWK.* Herb-herb combination for therapeutic enhancement and advancement: theory, practice and future perspectives. Molecules 2013; 18: 5125-5141

[2] *Zhao HY, Fan MX, Wu X, Wang HJ, Yang J, Si N, Bian BL.* Chemical profiling of the Chinese herb formula Xiao-Cheng-Qi decoction using liquid chromatography coupled with electrospray ionization mass spectrometry. J Chromatogr Sci 2013; 51: 273-285

[3] *Su SL, Cui WX, Zhou W, Duan JA, Shang EX, Tang YP.* Chemical fingerprinting and quantitative constituent analysis of Siwu decoction categorized formulae by UPLC-QTOF/MS/MS and HPLC-DAD. Chin Med 2013; 8: 5

[4] *Yan GL, Zou D, Zhang AH, Tan YL, Sun H, Wang XJ.* UPLC-Q-TOF-MS/MS fingerprinting for rapid identification of the chemical constituents of Ermiao Wan. Analytical Methods 2015; 7: 846-862

[5] *Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriot P, Nicholls AW, Reily MD, Thaden JJ, Viant MR.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initative (MSI). Metabolomics 2007; 3: 211-221

[6] *Liu GQ, Dong J, Wang H, Hashi Y, Chen SH.* Characterization of alkaloids in *Sophora flavescens* Ait. by high-performance liquid chromatography-electrospray ionization tandem mass spectrometry. J Pharm and Biomed Anal 2011; 54: 1065-1072

[7] *Ma Y, Gao HM, Liu J, Chen LM, Zhang QW, Wang ZM.* Identification and determination of the chemical constituents in a herbal preparation, compound Kushen injection, by HPLC and LC-DAD-MS/MS J of Liq Chrom & Relat Tech 2014; 37: 207-220

[8] *Liu XJ, Cao MA, Li WH, Shen CS, Yan SQ, Yuan CS.* Alkaloids from *Sophora flavescens* Aition. Fitoterapia 2010; 81: 524-527

[9] *Zhang J, Chen Z.* Determination of Matrine and Oxymatrine in *Sophora Flavescens* by Nonaqueous Capillary Electrophoresis-Electrospray Ionization-Ion Trap-Mass Spectrometry. Anal lett 2013; 46: 651-662

[10] *Lin T, Shieh B.* Furoquinoline alkaloids from Pai-Shen-Pi (*Dictamnus dasycarpus* Turcz) in Rutaceae. Huaxue 1986; 44: 96-100

[11] *Guo LN, Pei YH, Chen G, Lu X, Xu H, Liu JC.* Three new compounds from *Dictamnus dasycarpus*. J Asian Nat Prod Res 2012; 14: 210-215

[12] *Lin CC, Wu CI, Lin TC, Sheu SJ.* Determination of 19 rhubarb constituents by high-performance liquid chromatography-ultraviolet-mass spectrometry. J Sep Sci 2006; 29: 2584-2593

[13] *Jin W, Wang YF, Ge RL, Shi HM, Jia CQ, Tu PF.* Simultaneous analysis of multiple bioactive constituents in *Rheum tanguticum* Maxim. ex Balf. by high-performance liquid chromatography coupled to tandem mass spectrometry. Rapid Commun Mass Spectrom 2007; 21: 2351-2360

[14] *Ng LK, Lafontaine P, Vanier M.* Characterization of cigarette tobacco by direct electrospray ionization-ion trap mass spectrometry (ESI-ITMS) analysis of the aqueous extract--a novel and simple approach. J Agric and Food Chem 2004; 52: 7251-7257

[15] *Deng JW, Yang YY.* Chemical fingerprint analysis for quality assessment and control of Bansha herbal tea using paper spray mass spectrometry. Anal Chim Acta 2013; 785: 82-90

[16] *Gu DY, Yang Y, Hang B, Lv QY, Aisa HA.* Characterization and Identification of the Chemical Compositions in a Traditional Uighur Medicine Prescription Yizhihao Granule by LC–ESI-QTOF-MS. J of Liq Chrom & Relat Tech 2015; 38: 229-242

[17] *Parveen I, Threadgill MD, Hauck B, Donnison I, Winters A.* Isolation, identification and quantitation of hydroxycinnamic acid conjugates, potential platform chemicals, in the leaves and stems of *Miscanthus x giganteus* using LC-ESI-MSn. Phytochemistry 2011; 72: 2376-2384

[18] *Clifford MN, Johnston KL, Knight S, Kuhnert N.* Hierarchical scheme for LC-MSn identification of chlorogenic acids. J Agricl and Food Chem 2003; 51: 2900-2911

[19] *Han J, Ye M, Xu M, Sun JH, Wang BR, Guo D.* Characterization of flavonoids in the traditional Chinese herbal medicine-Huangqin by liquid chromatography coupled with electrospray ionization mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2007; 848: 355-362

[20] *Zhang L, Zhang RW, Li Q, Lian JW, Liang J, Chen XH, Bi KS.* Development of the Fingerprints for the Quality Evaluation of *Scutellariae Radix* by HPLC-DAD and LC-MS-MS. Chromatographia 2007; 66: 13-20

Legends for Tables

Table **1** Putatively identified compounds in the HHCF by LC-MS/MS in positive ionization mode.

Table **2** Putatively identified compounds in the HHCF by LC-MS/MS in negative ionization mode.

Legends for Figures

Fig. **1** Analytical platform and concept of the developed data mining method.

Fig. **2** Fragmentation pathways of the five most abundant compounds putatively identified in positive (P) and negative (N) ionization mode.

Table 1 Putatively identified compounds in the HHCF by LC-MS/MS in positive ionization mode.

R_t (min)	No.	*m/z*	Adduct ion(s)	MS/MS fragment ions (*m/z*)	Source	Molecular formula	Identity
3.36	**P1**	191	$[M+H]^+$	162, 148	SOP	$C_{11}H_{14}N_2O$ (190.24)	Cytisine
3.36	**P2**	196	$[M+NH_4]^+$	153, 137, 119, 109	PHE	$C_9H_6O_4$ (178.14)	Aesculetin
	P3	205	$[M+H]^+$	162, 146, 108	SOP	$C_{12}H_{16}N_2O$ (204.27)	N-Methylcytisine
	P4	215	$[M+2Na-H]^+$	171, 125	RHE	$C_7H_6O_5$ (170.12)	Gallic acid
3.75	**P5**	261	$[M+H]^+$	243, 164, 146, 114	SOP	$C_{15}H_{20}N_2O_2$ (260.33)	Baptifoline
	P6	265	$[M+H]^+$	247, 150, 136, 112	SOP	$C_{15}H_{24}N_2O_2$ (264.36)	5α-Hydroxymatrine OR 9α-Hydroxymatrine
4.15	**P7**	265	$[M+H]^+$	247, 150, 136, 127, 112	SOP	$C_{15}H_{24}N_2O_2$ (264.36)	14β-Hydroxymatrine
4.40	**P8**	180	$[M+H]^+$	151, 121, 103	PHE	$C_{11}H_{18}NO$ (180.27)	Candicine
4.75	**P9**	245	$[M+H]^+$	199, 162, 148, 122	SOP	$C_{15}H_{20}N_2O$ (244.33)	Anagyrine
	P10	265	$[M+H]^+$	247, 150, 136, 112	SOP	$C_{15}H_{24}N_2O_2$ (264.36)	5α-Hydroxymatrine OR 9α-Hydroxymatrine
5.00	**P11**	247	$[M+H]^+$	176, 150, 136	SOP	$C_{15}H_{22}N_2O$ (246.35)	5,6-Dehydrolupanine
5.52	**P12**	263	$[M+H]^+$	245, 166, 150, 128	SOP	$C_{15}H_{22}N_2O_2$ (262.35)	Mamanine
5.72	**P13**	192	$[M+H]^+$	177, 163, 149, 119	PHE	$C_{10}H_9NO_3$ (191.18)	Noroxyhydrastinine
	P14	249	$[M+H]^+$	176, 150, 112	SOP	$C_{15}H_{24}N_2O$ (248.36)	Allomatrine OR Isomatrine OR Matrine OR Sophoridine
		519	$[2M+Na]^+$	271			
		497	$[2M+H]^+$	249, 150			
		271	$[M+Na]^+$	No fragment ion			
6.47	**P15**	247	$[M+H]^+$	227, 179, 150, 136	SOP	$C_{15}H_{22}N_2O$ (246.35)	Sophocarpine OR Isosophocarpine
	P16	249	$[M+H]^+$	176, 150, 112	SOP	$C_{15}H_{24}N_2O$ (248.36)	Allomatrine OR Isomatrine OR Matrine OR Sophoridine
		519	$[2M+Na]^+$	271			
		497	$[2M+H]^+$	248, 249			
8.02	**P17**	263	$[M+H]^+$	245, 195, 150, 136	SOP	$C_{15}H_{22}N_2O_2$ (262.35)	Oxysophocarpine
		525	$[2M+H]^+$	263, 245, 150, 136			
	P18	265	$[M+H]^+$	247, 205, 150, 136, 112	SOP	$C_{15}H_{24}N_2O_2$ (264.36)	Oxymatrine OR Oxysophoridine
		529	$[2M+H]^+$	265, 247, 205, 136			
	P19	266	$[M+NH_4]^+$	249, 248, 205, 176, 150, 112	SOP	$C_{15}H_{24}N_2O$ (248.36)	Lupanine

Table 1 Putatively identified compounds in the HHCF by LC-MS/MS in positive ionization mode.

R_t (min)	No.	*m/z*	Adduct ion(s)	MS/MS fragment ions (*m/z*)	Source	Molecular formula	Identity
10.42	**P20**	247	$[M+H]^+$	179, 176, 150, 136	SOP	$C_{15}H_{22}N_2O$ (246.35)	(+)-7,11-Dehydromatrine
11.30	**P21**	263	$[M+H]^+$	245, 235, 195, 112	SOP	$C_{15}H_{22}N_2O_2$ (262.35)	9α-Hydroxysophocarpine
11.48	**P22**	263	$[M+H]^+$	245, 218, 164, 150	SOP	$C_{15}H_{22}N_2O_2$ (262.35)	Leontalbinine N-oxide
		525	$[2M+H]^+$	NA			
	P23	265	$[M+H]^+$	247, 205, 150, 136, 112	SOP	$C_{15}H_{24}N_2O_2$ (264.36)	Oxymatrine OR Oxysophoridine
		529	$[2M+H]^+$	NA			
11.88	**P24**	243	$[M+H]^+$ OR $[M]^+$	No fragment ions	DIC	$C_{12}H_{18}O_5$ (242.27)	Dasycarpusenester A
						$C_{14}H_{13}NO_3$ (243.26)	O-Ethylnor-γ-fagarine
20.08	**P25**	342	$[M+H]^+$	192, 177	PHE	$C_{20}H_{23}NO_4$ (341.40)	Tetrahydrojatrorrhizine
21.12	**P26**	344	$[M+NH_4]^+$	193, 165, 107	RHE	$C_{16}H_{22}O_7$ (326.34)	4-(4'-Hydroxylphenyl)-2-butanone 4'-O-β-D-glucoside
	P27	349	$[M+NH_4]^+$	291, 247, 185, 127	RHE	$C_{15}H_{14}O_6$ (290.27)	Epicatechin
22.08	**P28**	344	$[M+H]^+$	299, 175, 143	PHE	$C_{15}H_{25}N_3O_6$ (343.37)	Methyl-N-{[(2-methyl-2-propanyl)oxy]carbonyl}glycylprolylglycinate
22.91	**P29**	342	$[M+H]^+$	297, 265, 192, 130, 116	PHE	$C_{15}H_{23}N_3O_6$ (341.36)	2-{3,3-Bis[(2-hydroxyethyl)amino]-2-nitroprop-2-en-1-ylidene}-5,5-dimethylcyclohexane-1,3-dione
24.33	**P30**	592	$[2M+NH_4+H]^+$	286, 163	PHE	$C_{18}H_{13}N_3O$ (287.32)	Rutecarpine
26.84	**P31**	314	$[M+H]^+$	269, 137, 115, 107	PHE	$C_{19}H_{23}NO_3$ (313.39)	Evoeuropine
31.32	**P32**	356	$[M+H]^+$	311, 247	PHE	$C_{13}H_{21}N_7O_5$ (355.35)	2-Amino-N-(3-amino-2-hydroxypropyl)adenosine
52.87	**P33**	366	$[M]^+$	321, 306, 292, 278	PHE	$C_{20}H_{18}NO_4$ (336.36)	Berberine

Table 2 Putatively identified compounds in the LC-MS/MS in negative ionization mode.

Rt (min)	No.	*m/z*	Adduct ion(s)	MS/MS fragment ions (*m/z*)	Source	Molecular formula	Identity
2.34	**N1**	193	$[M-H]^-$	113, 103	SCU	$C_{10}H_{10}O_4$ (194.18)	Glucuronic acid
2.46	**N2**	191	$[M-H]^-$	173, 127, 111	PHE	$C_7H_{12}O_6$ (192.17)	Quinic acid
	N3	223	$[M-H]^-$	205, 129, 111	SOP	$C_{10}H_{14}N_2O_5$ (242.23)	Sinapiic acid
3.50	**N4**	191	$[M-H]^-$	155, 131, 129, 111	PHE	$C_6H_8O_7$ (192.12)	Citric acid
	N5	331	$[M-H]^-$	271, 241, 211, 169, 125	RHE	$C_{13}H_{16}O_{10}$ (332.26)	Galloylglucose (i.e. 1-*O*-Galloyl-β-D-glucose or 6-*O*-Galloyl-β-D-glucose) OR Glucopyranosyloxyl gallic acid (i.e. Gallic acid-3-O-β-D-glucoside or Gallic acid-4-O-β-D-glucoside)
4.08	**N6**	331	$[M-H]^-$	271, 241, 211, 169, 125	RHE	$C_{13}H_{16}O_{10}$ (332.26)	
4.61	**N7**	331	$[M-H]^-$	271, 241, 211, 169, 125	RHE	$C_{13}H_{16}O_{10}$ (332.26)	
5.35	**N8**	125	$[M-H]^-$	125	RHE	$C_6H_6O_3$ (126.11)	Pyrogallol
	N9	169	$[M-H]^-$	125, 107	RHE	$C_7H_6O_5$ (170.12)	Gallic acid
	N10	331	$[M-H]^-$	271, 241, 211, 169, 125	RHE	$C_{13}H_{16}O_{10}$ (332.26)	Galloylglucose (i.e. 1-*O*-Galloyl-β-D-glucose or 6-*O*-Galloyl-β-D-glucose) OR Glucopyranosyloxyl gallic acid (i.e. Gallic acid-3-O-β-D-glucoside or Gallic acid-4-O-β-D-glucoside)
8.71	**N11**	255	$[M-H]^-$	193, 179, 165, 107	SOP	$C_{11}H_{12}O_7$ (256.21)	Piscidic acid
		511	$[2M-H]^-$				
13.88	**N12**	577	$[M-H]^-$	425, 407, 289, 245, 125	RHE	$C_{30}H_{26}O_{12}$ (578.52)	Procyanidin B (Catechin dimers)
16.48	**N13**	289	$[M-H]^-$	245, 205, 151, 125, 109	RHE	$C_{15}H_{14}O_6$ (290.27)	Catechin
		579	$[2M-H]^-$	289, 245, 205, 151, 125			
	N14	353	$[M-H]^-$	205, 191, 179, 127, 109	PHE	$C_{16}H_{18}O_9$ (354.31)	Chlorogenic acid
		707	$[2M-H]^-$	NA			
18.15	**N15**	367	$[M-H]^-$	193, 191, 173, 149	PHE	$C_{17}H_{20}O_9$ (368.34)	3-*O*-Feruloylquinic acid
21.10	**N16**	325	$[M-H]^-$	NA	RHE	$C_{16}H_{22}O_7$ (326.34)	4-(4'-Hydroxylphenyl)-2-butanone 4'-O-β-D-glucoside
		371	$[M+FA-H]^-$	163, 121, 101			
		651	$[2M-H]^-$	NA			
	N17	415	$[M+Na-2H]^-$	NA	RHE	$C_{19}H_{22}O_9$ (394.37)	6-Hydroxymusizin-8-O- β-D-glucoside
		439	$[M+FA-H]^-$	393, 231, 113			
22.77	**N18**	289	$[M-H]^-$	245, 205, 151, 125, 109	RHE	$C_{15}H_{14}O_6$ (290.27)	Epicatechin
	N19	337	$[M-H]^-$	191, 173, 163, 111	PHE	$C_{20}H_{20}NO_4$ (338.38)	*p*-coumaroylquinic acid (not previously reported in Phellodendron)
24.72	**N20**	303	$[M-H]^-$	177, 151, 125	SCU	$C_{15}H_{12}O_7$ (304.25)	2',3,5,6',7-Pentahydroxyflavanone (Ganhuangemin)
26.26	**N21**	367	$[M-H]^-$	193, 191, 173	PHE	$C_{17}H_{20}O_9$ (368.34)	5-*O*-Feruloylquinic acid

26.65	**N22**	389	$[M-H]^-$	NA	RHE	$C_{20}H_{22}O_8$ (390.38)	Resveratrol-4'-O-β-D-glucoside OR Resveratrol 3-O-β-glucoside (Pieceid)
		435	$[M+FA-H]^-$	389, 227, 191, 185			
		779	$[2M-H]^-$	NA			
		825	$[2M+FA-H]^-$	NA			
	N23	503	$[M-H]^-$	191, 111	RHE	$C_{18}H_{32}O_{16}$ (504.44)	Melezitose
		584	$[M+Br]^-$	NA			
28.16	**N24**	229	$[M-H]^-$	185, 139, 117	PHE	$C_{14}H_{14}O_3$ (230.26)	Osthenol
32.06	**N25**	301	$[M-H]^-$	195, 173, 149, 125	SCU	$C_{15}H_{10}O_7$ (302.24)	3,5,7,2',6'-Pentahydroxyflavone (Viscidulin I)
34.57	**N26**	441	$[M-H]^-$	289, 271, 169, 125	RHE	$C_{22}H_{18}O_{10}$ (442.37)	Epicatechin 3-O-gallate
	N27	477	$[M-H]^-$	313, 169, 125	RHE	$C_{23}H_{26}O_{11}$ (478.45)	Isolindleyin
	N28	555	$[M-H]^-$	NA	RHE	$C_{28}H_{28}O_{12}$ (556.51)	Desoxyrhaponticin-6"-O-gallate
		591	$[M+Cl]^-$	477, 137			
	N29	547	$[M-H]^-$	457, 427, 367, 337	SCU	$C_{26}H_{28}O_{13}$ (548.49)	Chrysin-6-C-arabinosyl-8-C-glucoside
36.56	**N30**	477	$[M-H]^-$	313, 169, 125	RHE	$C_{23}H_{26}O_{11}$ (478.45)	Lindleyin
	N31	545	$[M-H]^-$	NA	RHE	$C_{25}H_{22}O_{14}$ (546.43)	Rhein-8-O-D-[6'-O-(3"-methoxylmalonyl)] glucoside
		591	$[M+FA-H]^-$	505, 477, 313			
		955	$[2M-H]^-$	477, 313			
37.25	**N32**	547	$[2M-H]^-$	457, 427, 367, 337	SCU	$C_{26}H_{28}O_{13}$ (548.49	Chrysin-6-C-glucosyl-8-C-arabonoside
38.43	**N33**	541	$[M-H]^-$	313, 227, 169	RHE	$C_{27}H_{26}O_{12}$ (542.49)	Resveratrol-4'-O-β-D-(2"-O-galloyl) glucoside
39.20	**N34**	541	$[M-H]^-$	313, 227, 169		$C_{27}H_{26}O_{12}$ (542.49)	Resveratrol-4'-O-β-D-(6"-O-galloyl) glucoside
45.08	**N35**	301	$[M-H]^-$	273, 269, 165, 139	SCU	$C_{16}H_{14}O_6$ (302.28)	4',5,7-trihydroxy-6-methoxyflavanone OR (2S)-7,2',6'-trihydroxy-5-methoxyflavanone
46.91	**N36**	431	$[M-H]^-$	311, 293, 225	RHE	$C_{21}H_{20}O_{10}$ (432.38)	Emodin-1-O-β-D-glucoside OR Emodin-8-O-β-D-glucoside OR Aloe-emodin 8-O-β-D-glucoside OR Aloe-emodin-3-CH_2-O-β-D-glucoside.
55.48	**N37**	481	$[M+Cl]^-$	NA	SOP	$C_{22}H_{22}O_{10}$ (446.40)	(-)-Maackiain-3-O-glucoside (Trifolirhizin)
		483	$[M+K-2H]^-$	NA			
		491	$[M+FA-H]^-$	283, 255			
58.94	**N38**	431	$[M-H]^-$	311, 269, 225	RHE	$C_{21}H_{20}O_{10}$ (432.38)	Emodin-1-O-β-D-glucoside OR Emodin-8-O-β-D-glucoside OR Aloe-emodin 8-O-β-D-glucoside OR Aloe-emodin-3-CH_2-O-β-D-glycoside.
63.95	**N39**	233	$[M-H]^-$	191, 175, 147	RHE	$C_{12}H_{10}O_5$ (234.20)	(5Z)-6-Hydroxy-3,4-dioxo-6-phenyl-5-hexenoic acid
71.02	**N40**	269	$[M-H]^-$	251, 223, 169,	SCU	$C_{15}H_{10}O_5$ (270.24)	5,6,7-Trihydroxyflavone (Baicalein) OR 5,7,8-Trihydroxyflavone (Norwogonin)

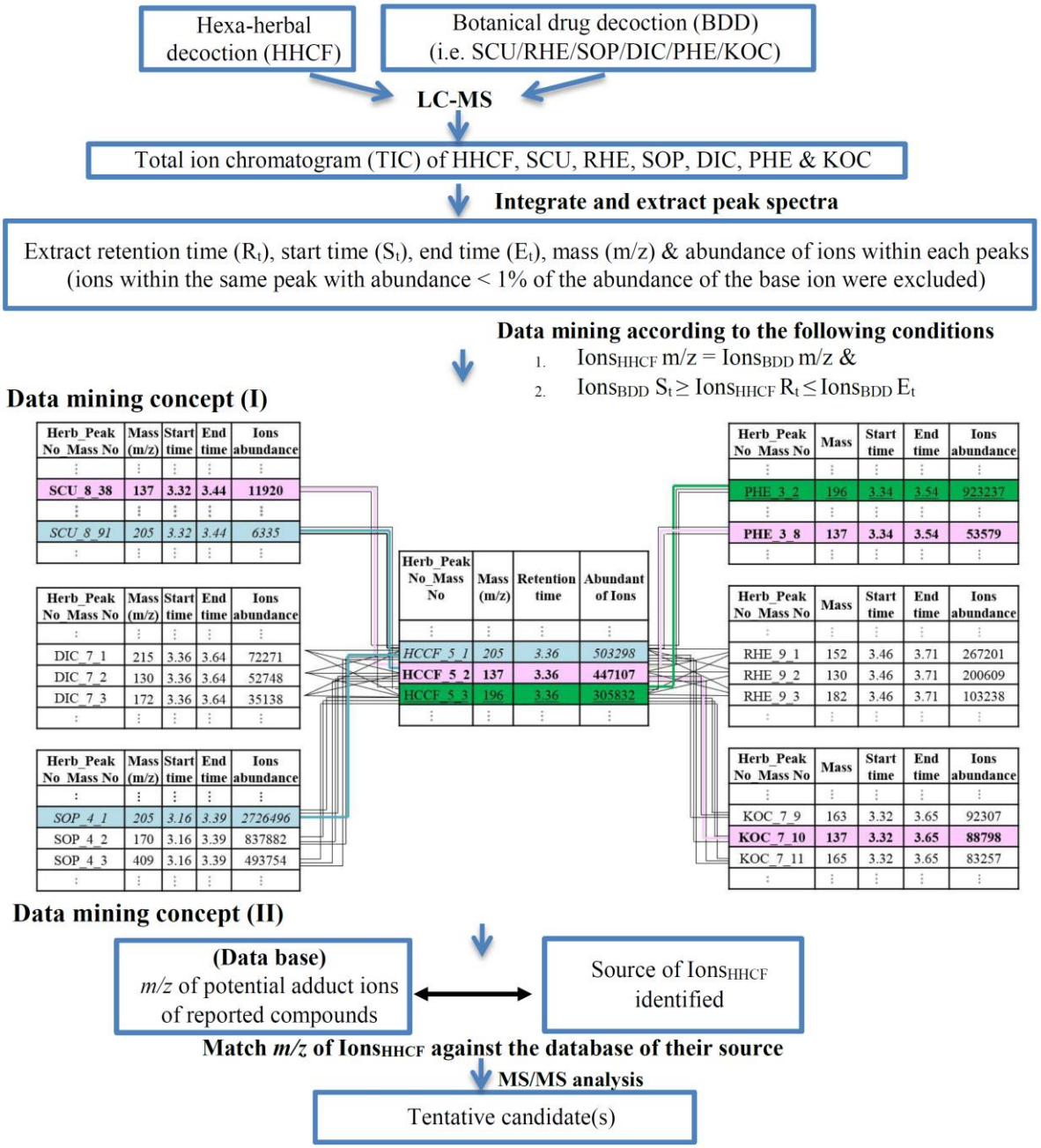

Fig. 1 Analytical platform and concept of the developed data mining method.

Data mining concept (I), three cases were encountered during data mining process i.e. 1) (Green cells) If an ion was present in both the HHCF and one of the botanical drugs and was eluted at the same time, this ion was assigned to be sourced from that particular botanical drug, 2) If an ion was present in the HHCF and more than one botanical drugs with same elution time; and a) (**Pink cells**) the abundance of the ion in the botanical drugs were similar These ions could be background noise, common herbal chemical components or fragments and were excluded (excluded ions are listed in Table S1 and required further study to assign their identity), b) (*Blue cells*) the abundance of the ion in the botanical drugs were dissimilar i.e. the abundance of the ion in one of the botanical drugs was at least two times the average of the abundances of the ion in botanical drugs that the ion was present in, the source of the ion was then considered to be the botanical drug that the ion was present at the highest abundance.

Data mining concept (II), after identifying the source of ions in the HHCF, ions were sorted by abundance in descending order. Ions with intensity> 1.5% of the ion that was present in the greatest abundance were subjected to MS/MS analysis. The in-house developed template will then carry out matching of the mass (*m/z*) of ions detected in the HHCF against the mass (of potential adduct ions) of reported compounds from their source. A mass database was developed for each botanical drug based on literature review and each involved mass of potential adduct ions (i.e. $[M+H]^+$, $[M]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, $[M+2Na-H]^+$, $[M+2K-H]^+$, $[2M+H]^+$, $[2M+Na]^+$, $[2M+NH_4]^+$, $[M-H]^-$, $[2M-H]^-$, $[M+Cl]^-$, $[M+Br]^-$, $[M+Na-2H]^-$, $[M+K-2H]^-$, $[M+FA-H]^-$, $[2M+FA-H]^-$, $[3M-H]^-$, $[M-H_2O-H]^-$) of reported compounds.

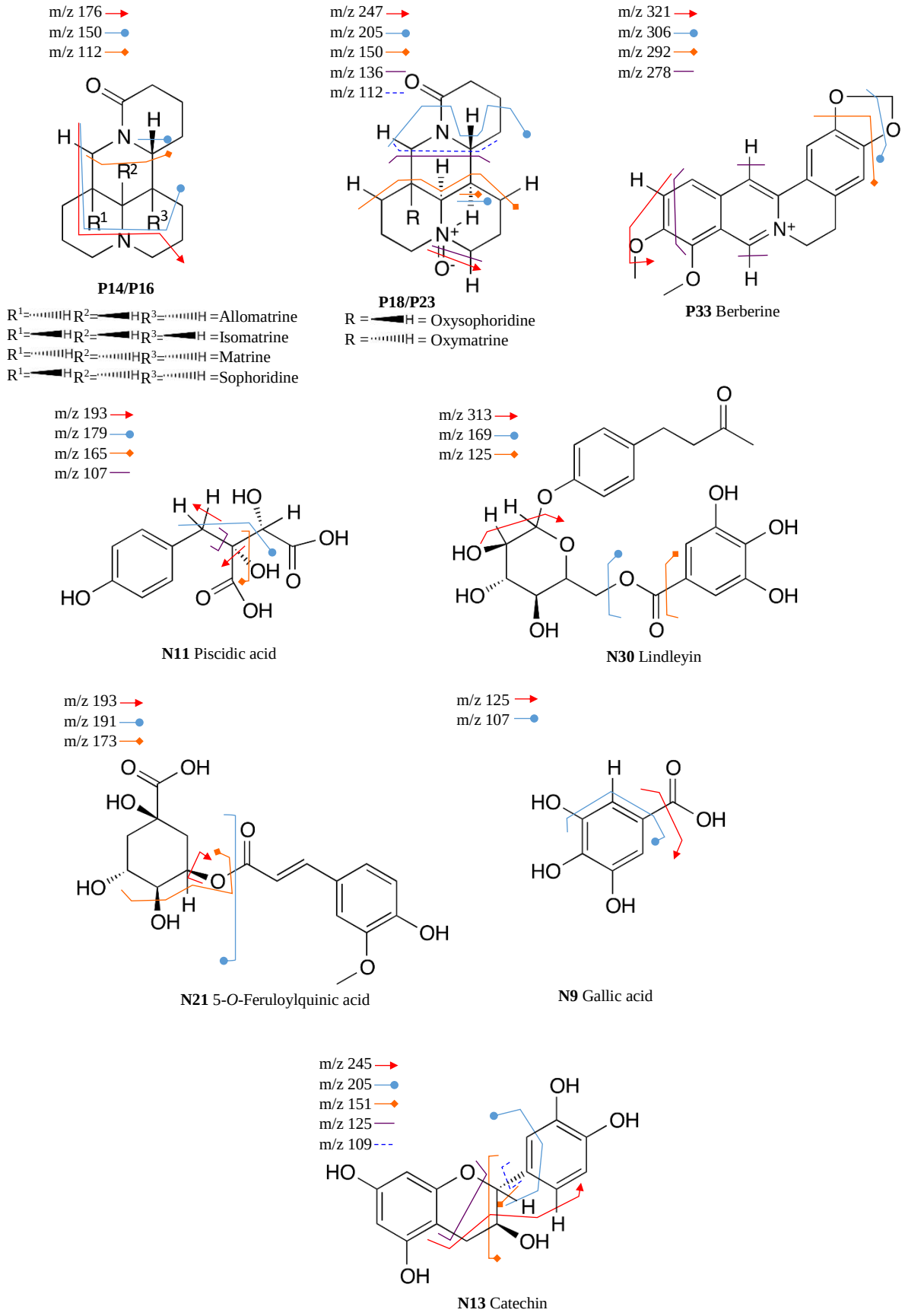

Fig. 2 Fragmentation pathways of the five most abundant compounds putatively identified in positive (P) and negative (N) ionization mode.