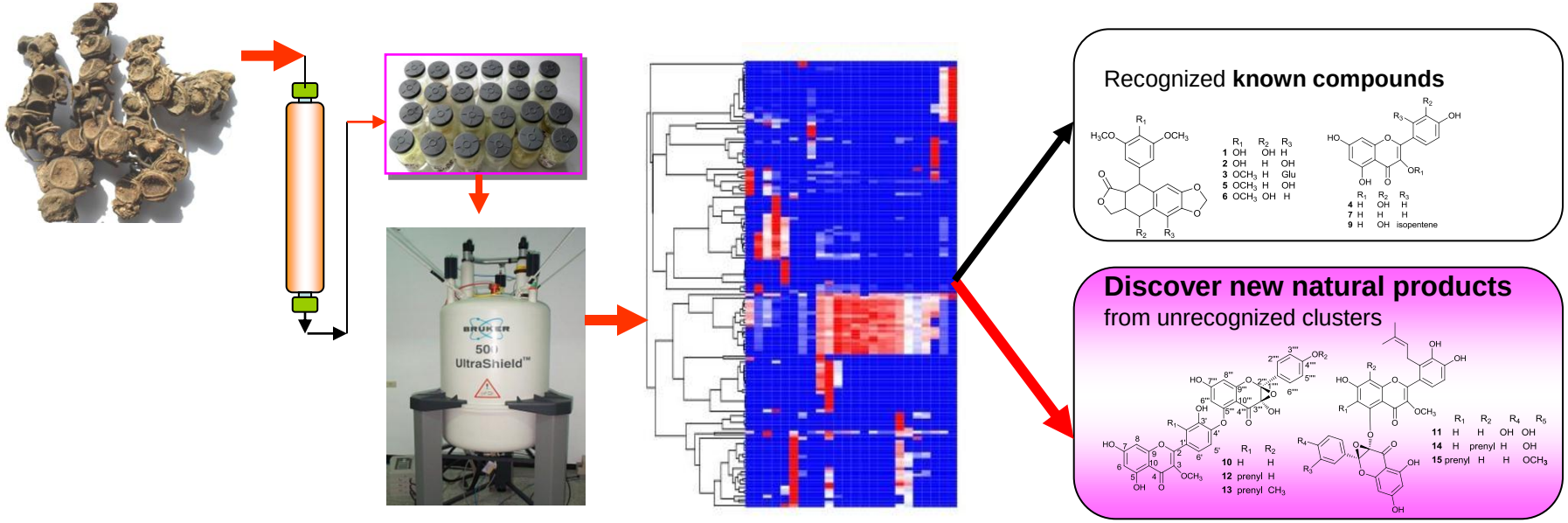
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



Off-line RPLC-^{13}C NMR HCA cluster

Journal Name

ARTICLE

A novel strategy for screening new natural products by combination of reversed-phase liquid chromatography fractionation and ^{13}C NMR pattern recognition: the discovery of new anti-cancer flavone dimers from *Dysosma versipellis* (Hance)

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Zhi Yang[a], Youqian Wu[a], Hui Zhou[b], Xiaoji Cao[c], Xin-hang Jiang[d], Kuiwu Wang[e], Shihua Wu[a,*]

Natural products have been a rich source for drug discovery. However, the traditional process of discovering new bioactive natural products is generally labor-intensive and time-consuming, and known natural products are frequently rediscovered. In this work, we presented a new screening strategy for the discovery of new natural products by combination of reversed-phase liquid chromatography (RPLC) and ^{13}C NMR pattern recognition. The known compounds were first recognized by ^{13}C NMR clustering analyses and on-line ^{13}C NMR database matching. The unrecognized ^{13}C NMR clusters and HPLC peaks were possibly new natural products and then further subjected to targeted isolation and purification for structural elucidation. Thus, this method may win a higher hit rate of new natural products than traditional process. As an example, we analyzed a cytotoxic sample extracted from roots of *Dysosma versipellis* (Hance) by RPLC fractionation and following ^{13}C NMR clustering analyses. As a result, 7 ^{13}C NMR clusters were recognized as 7 known compounds including 5 podophyllotoxins and 2 flavones corresponding 7 HPLC peaks by comparison with reported NMR data. 1 unrecognized ^{13}C NMR block including at least three unrecognized NMR clusters gave us clues for new natural products, guiding the following targeted isolation and purification, which resulted in the discovery of 6 new flavone dimers podoverine D, E, F, G, H and I together with the known podoverine A. Interestingly, these new flavone dimers expressed potential cytotoxicity to several cancer cells *in vitro*. To the best of our knowledge, this is the first document to demonstrate RPLC fractionation-^{13}C NMR pattern recognition strategy to rapidly discovery new natural products. It is an important advancement for natural product identification and metabolomic analyses.

1. Introduction

Natural products have been used as the most important resources for new chemical entities despite some combination chemistry-based methods may compose a large number of chemical entities or compounds library.[1-4] Natural products represent diversity of chemical structures and potent biological activities. There are many successful examples for new drug developments from natural products, especially for anti-cancer and anti-inflammatory drugs as well as antibiotics.[5-9] Therefore, screening new entities from natural products has been thought to be an efficient method to hit more potent drugs against drug resistance and disease evolution.

However, the extracts from natural resources like plants, animals or macromycetes are very complex. They usually contain complex chemical components and possess versatile chemical structures. Although a number of mass spectrometry (MS)[10-12] or NMR[13]-based methods coupled with on-line or off-line chromatographic separation have been developed for rapid characterization of natural products, the discovery of new natural products from complex natural resource is still a hard work in the natural product researches.[14] The traditional process of discovering new bioactive natural products is generally long and laborious, and known natural products are frequently rediscovered.[15] Even under bio-guided isolation, it is very possible that the obtained components are still known natural products because of the same or similar metabolism pathways found in the same or different natural product sources. Thus even in some un-explored plant materials, there are many known metabolites[16] Therefore, the purpose of this work is to establish an efficient strategy for the discovery of new natural products from a natural source without or with pre-separation and purification as little as possible.

As is well known, liquid chromatography (LC) is one of the most important separation techniques for isolation and purification of natural products. Some hyphenated techniques such as on-line

[a] Research Center of Siyuan Natural Pharmacy and Biotoxicology, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
[b] Department of Pharmaceutical Analysis and Drug Metabolism, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang Province 310058, China
[c] Research Center of Analysis and Measurement, Zhejiang University of Technology, 18 Chaowang Rd, Hangzhou, Zhejiang 310014, China
[d] Equipment & Technology Service Platform, College of Life Sciences, Zhejiang University, Hangzhou 310058, P.R. China
[e] School of Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou 310035, China
*Corresponding author. Tel./fax: +86-571-88206287; E-mail: drwushihua@zju.edu.cn
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

LC-NMR and LC-NMR-MS have been proved to be rapid and efficient methods to identify many components from natural complexities[17, 18] In most cases, however, off-line LC-NMR is still one of the major research modes in a natural product lab due to the expensive instruments and high costs of on-line operations.[19] In addition, it is easier to enrich minor components by large-scale sample injection using off-line LC preparation than that of expensive and small-scale on-line LC-NMR analyses.

It is well known that NMR has been used as an efficient and unambiguous method to determine the chemical structures of natural products for a long time[13] It can be recorded from crude, even one-step extracts without derivatization or purification, allowing the simultaneous detection of diverse groups of secondary metabolites (flavonoids, alkaloids, terpenoids, and so on).[20, 21] There are two major NMR spectra, one is the ^{1}H NMR spectrum and the other is the ^{13}C NMR spectrum. In ^{1}H NMR spectrum, signals are proportional to their molar concentration, which makes a direct comparison of concentrations of all compounds possible without the need for calibration curves of each individual compound.[22] By contrast, in ^{13}C NMR spectrum there is wider window range (0-250 ppm but for ^{1}H NMR only at 0-15 ppm) and less overlapping signals, which make it more suitable for dereplication and comparison. In addition, there are many available ^{13}C NMR databases for known compounds, such as Human Metabolome Database,[23] the Yeast Metabolome Database,[24] MicroNMR database[25] and Natural Products ^{13}C NMR Database,[26] which may help to elucidate the structures of known compounds rapidly once the ^{13}C NMR shifts were obtained.

During the last twenty-five years, there are numbers of Omics methods such as genomics, transcriptomics, proteomics and metabolomics developed for analyses of the complex biological samples. Due to the advantages in quantitative and qualitative structural characterization described above, NMR has also been widely used for comprehensive metabonomic or metabolomic analyses ranging from pattern recognition to biological interpretation since ten years ago.[27-29] Using the pattern recognition techniques of Omics, ^{1}H NMR signals of a compound in a complex sample can be extracted for structural identification. Therefore, ^{1}H NMR–based metabolomic or metabonomic method has been widely used for the determination of systemic biochemical profiles and research of the regulation of function in whole organisms by analyzing biofluids and tissues.[30, 31] Recently, ^{13}C NMR pattern recognition technique has also been developed,[32-34] especially with the invention of cryogenic probe technology, compensating for the inherently low sensitivity of natural abundance ^{13}C NMR spectroscopy.[21, 35] More recently, Dr. Jan Hubers and her collaborators have demonstrated a successful centrifugal partition chromatography (CPC) separation-^{13}C NMR pattern recognition method to identify the major components of several plants including *Anogeissus leiocarpus* Guill and lichen *Pseudevernia furfuracea* without further fractionation and purification.[36, 37] It seems very efficient for identification of the known compounds among the prominent components from the complex extracts.

Therefore, in this work we established a strategy to distinguish the known and new natural products by combination of RPLC fractionation and ^{13}C NMR pattern recognition technique. As an example, the ethanol extract of *Dysosma versipellis* (Hance) was selected, which was found to have potent cytostatic activities (below 1 μg/mL) against several cancer cells *in vitro,* such as human breast cancer Bcap37, human hepatoma HepG2, doxorubicin-resistant human hepatoma R-HepG2, mouse melanoma B16 and brain glioma GL261 cells in our course of screening new anti-cancer natural products. Although a number of podophyllotoxins and flavonoids have been isolated and identified from this plant and other plant of podophyllum taxa,[38-42] there are still some un-identified components by on-line HPLC-MS^{n} analysis.[43-45] Thus, we selected the extract as the sample. As expected, whole results demonstrated that this strategy by combination of RPLC fractionation and ^{13}C NMR pattern recognition was very efficient in natural product research, not only for recognition of the known metabolites without standards but also for dereplication and search for new natural products.

2. Experimental

2.1. Reagent and materials

Organic solvents for the chromatographic fractionation separation, including petroleum ether, ethyl acetate and methanol were of analytical grades (Sinopharm Chemical Reagent Co., Shanghai, China). Methanol used for HPLC was of chromatographic grade (Merck, Darmstadt, Germany). The water was purified by means of a water purifier (18.2 MΩ, Wanjie Water Treatment Equipment, Hangzhou, China). Silica gel (200-300 mesh) used for flash chromatography (40 cm length and 14.5 cm id) was obtained from Qingdao Haiyang Chemical Co, Qingdao, China. The spherical silica gel (ODS-C_8, 50 μm) was purchased from Beijing Greenherbs Science and Technology Development Co. Beijing, China.

The dry roots of *D. versipellis* (Hance) were bought from a drug market in Bozhou (Anhui, China). The species was identified by the Institutes of Plant Sciences, College of Life Sciences, Zhejiang University, China.

2.2. Preparation of crude sample

The dry roots of *D. versipellis* (Hance) were ground to a fine powder. The power (2 Kg) was then extracted three times with 4 L of 95% ethanol at boiling point, each time for 2 h. All extracts were combined and evaporated to dryness at 45°C under reduced pressure, and 420 g residue was finally obtained. Then the residue was subjected to a flash glass silica chromatography (40 cm length and 14.5 cm i.d.). After elution with 10 L of petroleum ether-ethyl acetate (1:1, v/v), all effluents were collected as the sample for RPLC separation.

2.3. RPLC fractionation

The RPLC fractionation was performed on a glass column (120 cm length and 2.5 cm id) packed with the spherical silica gel (ODS-C_8, 50 μm). The preparative chromatography system was composed of two P270 pumps and an on-line solvent-gradient mixer (Elite Analytical Instrument Co., Ltd., Dalian, China), a six-port valve with a 20 mL sample loop, a UV 230^{+} spectrometer (Elite Analytical Instrument Co., Ltd., Dalian, China), a BSZ-100 fraction collector and an EC2000 ChemStation (Elite Analytical Instrument Co., Ltd.,

Dalian, China). 5 g sample of *D. versipellis* was first loaded on the top of the column and then eluted with methanol (solvent A) and 0.5% acetic acid aqueous solution (solvent B) in a gradient elution mode as follows: 0-60 min, A from 10% to 40%, B from 90% to 60%; 60-240 min, A from 40% to 80%, B from 60% to 20%; 240-300 min, A from 80% to 100%, B from 20% to 0%. The flow rate was 8 mL/min, consuming 1.4 L methanol and 1 L water during the whole elution process. The effluents were collected every two minutes and monitored at 254 nm all the time. The effluents with similar composition were combined, evaporated to dryness and lyophilized. Finally 23 fractions (S1-S23) were collected.

2.4. HPLC analysis of the crude sample and RPLC fractions

HPLC analysis was performed on an Agilent 1100 system, equipped with a G1379A degasser, a G1311A Quat Pump, a G1367A Wpals, a G1316A column oven, a G1315B diode assay detector and an Agilent ChemStation. The column used was a reversed phase column (Zorbax SB-$C_{18,}$ 250 mm × 4.6 mm id., 5μm). The flow rate was 0.8 mL/min and the column temperature was 30°C. The injection volume was 10 μL. Methanol-0.1% TFA aqueous solution system was used as the mobile phases in a gradient mode as follows: 0-5 min, methanol from 10% to 30%; 5-35 min, methanol from 30% to 70%; 35-45 min, methanol from 70% to 100%; 45-50 min, methanol from 100% to 10%; 50-55 min, methanol was maintained at 10%. The effluent was monitored by a diode assay detector at multiple wavelengths, including 254 nm, 338 nm, 210 nm, 230 nm and 280 nm.

2.5. NMR analysis and pattern recognition of ^{13}C chemical shifts

All samples were analyzed with the same acquisition and processing parameters. Every fraction of 20 mg was dissolved in 500 μL DMSO-*d*6. NMR analysis was performed at 298K on a Bruker DMX-500 spectrometer. The ^{13}C NMR spectra were acquired at 125 MHz. A standard zgpg pulse sequence was used with an acquisition time of 0.865s and a relaxation delay of 1.5s. For each sample, 1448 scans were co-added to obtain a satisfactory signal-to-noise (S/N) ratio. The spectral width was 300 ppm ranging from -40 ppm to 260 ppm. The spectra data were handled with the ACD/NMR processor Academic Edition (ACD/Labs Release 12.0, Ontario, Canada). The central resonance of DMSO-*d*6 was calibrated at δ 39.48 ppm and its intensity was set as 1.

Positive ^{13}C peak signals were collected automatically with a minimum intensity threshold of 0.005 and exported to the professional software Similarity Evaluation System for Chromatographic Fingerprint of Traditional Chinese Medicine (Version 2004A) for calibration the ^{13}C NMR shifts of all fractions (signals of DMSO-*d6* were removed). The ^{13}C NMR spectrum of the crystallized podophyllotoxin (S24) was selected as the reference spectrum with a chemical shift window width of 0.2 ppm and then multi-point calibration and matching was performed automatically for normalization of the peaks.

The normalized ^{13}C NMR data were imported into GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/index.html, Broad Institute) for pattern recognition. Hierarchical clustering analysis was directly applied on raw ^{13}C peak intensity values. The classification was performed on the rows and columns. The Eucidian distance method was used to measure the proximity between samples and the One minus pearson correlation method was performed to agglomerate the ^{13}C NMR data. The resulting ^{13}C NMR chemical shifts clusters were visualized as dendrograms on a 2D heat map. The deeper the red colour in the map, the higher the relative intensity of ^{13}C peaks.

Each ^{13}C chemical shift cluster obtained from GENE-E was submitted to an on-line structure search engine of MICRONMR database (Shanghai Micronmr Infor Technology Co., Ltd.) for structural determination of new and known compounds. Up-to-now, MICRONMR database has collected the ^{13}C NMR data of more than 701,000 organic compounds and related information. A fuzzy search mode was employed, matching tolerance was set at 1 and the lowest similarity was set at 80%. It is possible to provide the exact and similar structures once ^{13}C NMR data were inputted.

2.6. Targeted isolation and structural identification of unrecognized NMR clusters

In order to validate the structures recognized by RPLC- ^{13}C NMR pattern recognition, the several fractions contained the targeted components were selected for purification of the targets. The structures of the purified components were identified by MS, MS^n, 1D and 2D NMR, and CD spectrum. MS analysis was performed on ThermoQuest Finnigan LCQDECA system equipped with an atmospheric ionization source (ThermoQuest LC–MS Division, San Jose, CA, USA). The ESI–MS^n spectra were acquired in positive or/and negative ion modes. The mass spectrometry detector (MSD) parameters were as follows: capillary temperature, 250°C; spray voltage, 4.5 kV; capillary voltage, -15 V in (-) ESI, 24 V in (+) ESI; lens voltage, 18 V in (-) ESI, -16 V in (+) ESI; sheath gas flow 60 arbitrary units of nitrogen; auxiliary gas flow 14 arbitrary units of nitrogen.

CD spectra were recorded on a Biologic MOS-450 AF-CD spectropolarimeter. The samples were dissolved in ethanol at concentrations of 0.1 mg/mL. The CD signal was recorded every 0.5 nm with a 0.5 s signal averaging for each point. Each spectrum was recorded twice and averaged.

2.7. Evaluation of cytotoxicity of new flavone dimers in vitro

The cytotoxicity of new flavones dimers was measured by 3-(4,5-dimethylthiazol-z-yl)-2,5-diphenyl tetrazolium bromide (MTT) (Sigma, MO, USA) assay based on the ability of live cells to cleave the tetrazolium ring to a molecule that absorbed at 570 nm on human breast cancer Bcap37, human hepatoma HepG2, doxorubicin-resistant human hepatoma R-HepG2, mouse melanoma B16 and brain glioma GL261 cells as reported process[46] with minor modification. In short, cells undergoing exponential growth were suspended in fresh medium at a concentration of 1×10^5 cells/mL and inoculated in a 96-well flat bottomed plate in a volume of 100 μL/well. Cells were stabilized by incubation for 24 h at 37°C and 100-μL aliquots of each drug sample were added to wells. The plate was incubated at 37°C for 48h. After 48 h, 20 μL MTT solution (5 mg/1 mL PBS) was added to each well, and the cells were further incubated at 37°C for 4 h. At the end of incubation period, the medium was discarded, washed with PBS for two times, 150 μL DMSO/well was added to 96-well plate to solubilize

formazan crystals, following by reading on a scanning multi-well spectrophotometer.

3. Results and discussion

3.1. HPLC analyses of crude sample

It has been known that there are more than 80 chemical components isolated from podophyllum taxa by various chromatographic methods.[41] The major active components of *D. versipellis* (Hance) are found to be podophyllotoxins and flavonoids.[40, 46] Previous studies showed that podophyllotoxins had stronger absorption at 254 nm[46] while flavonoids at about 330 nm.[47] Therefore, multiple-wavelength detection was set on DAD detector to monitor the HPLC elution peaks of *D. versipellis*. As shown in Fig. 1, there were more than 10 major components in the sample of *D. versipellis*. Most of podophyllotoxins were detected at 254 nm but not 338 nm. In contrast, most of flavonoids were found to be detected both at 254 nm and 338 nm in spite of the characteristic absorption was at 338 nm. Thus the further RPLC fractions were detected at 254 nm.

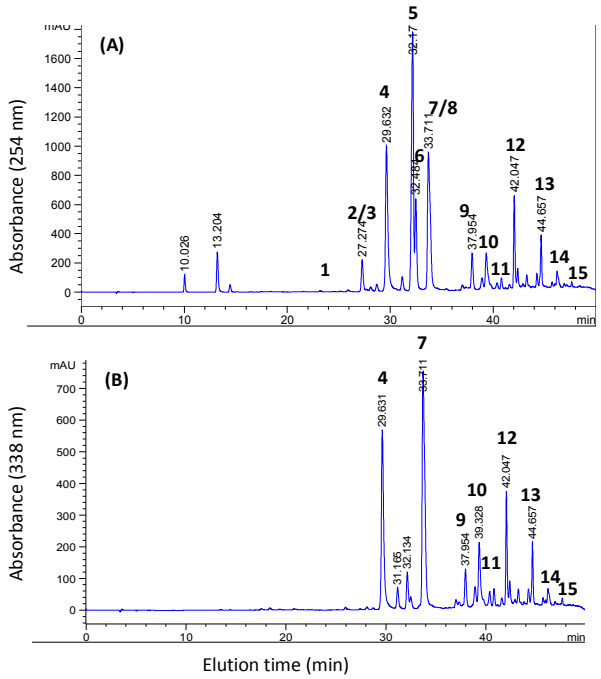


Fig. 1. HPLC analysis of the crude sample prepared from *Dysosma versipellis* (Hance) using a reversed-phase column (Zorbax SB-C_{18}, 250 mm length ×4.6 mm i.d., 5μm) with the detection wavelength of (A) 254nm and (B) 338 nm. The flow rate was 0.8 mL/min and the column temperature was 30°C. The injection volume was 10 μL. Methanol-0.1% TFA aqueous solution system was used as the mobile phases in a gradient mode as follows: 0-5 min, methanol from 10% to 30%; 5-35 min, methanol from 30% to 70%; 35-45 min, methanol from 70% to 100%; 45-50 min, methanol from 100% to 10%; 50-55 min, methanol was maintained at 10%.

3.2 RPLC fractionation of crude sample of *D. versipellis*

Although column chromatography using solid support matrix may result in some potential complications, such as irreversible adsorption (sample loss), tailing of solute peak, and contamination, it is still a simple and efficient method to enrich the component or fractionate the crude sample into several purer fractions rapidly.[48] Thus a RPLC column was used for fractionation of *D. versipellis*. As shown in Fig. 2, after RPLC fractionation, 23 fractions were obtained. Clearly, there were at least 1 to 3 major components in most of the fractions (Fig. 2). In addition, due to the relative low resolution of the RPLC separation, some component peaks, such as peaks 4, 5 and 7, were distributed widely in different fractions (Fig. 2). Table 1 listed the relative contents of 15 peaks found in the fractions.

It should be noted that although these fractions were not pure and contained several major components, some components in the fractions may be further purified by simple crystallization and re-crystallization. For example, podophyllotoxin (**6**) in the fraction S5 (Fig. 2) was easily crystallized and re-crystallized in 50% methanol aqueous solution. The purity of podophyllotoxin (**6**) was up to 98% after crystallization and re-crystallization while the purity was only 40% in fraction S5. In addition, in order to validate the following ^{13}C NMR-map based pattern recognition strategy, the crystallized podophyllotoxin was used as a sample (S24, Fig. 2) in the following ^{13}C NMR analyses

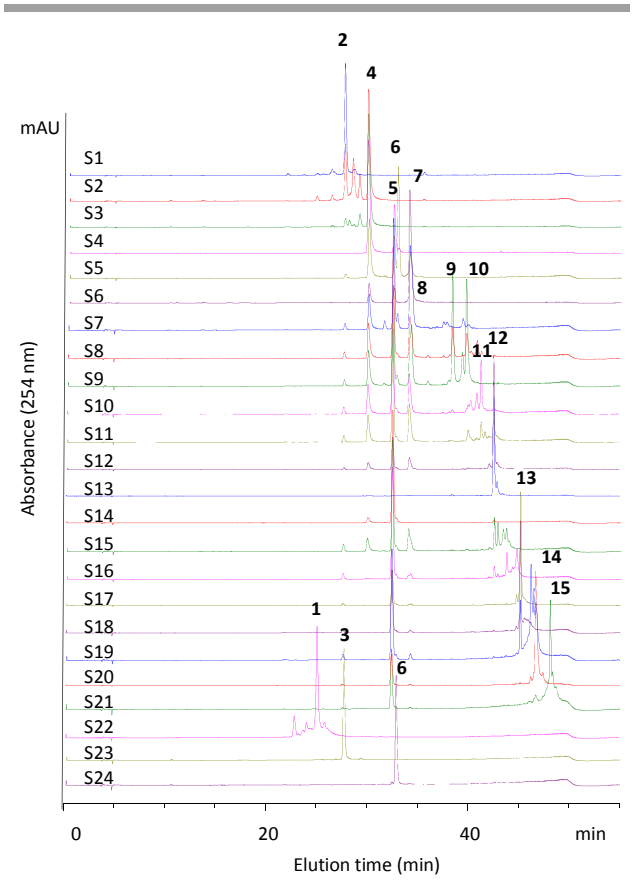


Fig. 2. Fingerprinting profiles of the RPLC fractions (S1-S23) and crystallized podophyllotoxin (S24) by HPLC. The HPLC conditions were the same as the Fig. 1

Table 1. The peak contents in the RPLC fractions.

Peak NO.	Retention time (min)	Relative content by HPLC analyses (% peak area)		The fraction for preparation	Identified name
		In the fractions	In the crude sample		
1	24.71	S22 (58.5%）	0.42	S22	4'-demethylpodophyllotoxin glucoside
2	27.18	S1 (72.7%), S2 (19.8%)	2.32*	S1	α-peltatin
3	27.97	S2 (19.6%), S23 (91.8%)	2.32*	S23	β-peltatin glucoside
4	29.48	S2 (46.3%), S3 (75.9%), S4 (70.4%), S5 (28.9%), S6 (12.1%), S7 (12.6%), S8 (14.1%), S9 (7.3%), S10 (10.2%), S11 (10.6%)	11.46	S3	quercetin
5	32.02	S4 (23%), S5 (14.5%), S6 (7.1%), S7 (28.9%), S8 (33%), S9 (15.5%), S10 (28.5%), S11 (33.5%), S12 (19.9%), S14 (87.6%), S15 (37%), S16 (42.45), S17 (21.45), S18 (31.3%), S19 (20.3%), S20(14.8%), S21(12.6%)	23.19	S14	β-peltatin
6	32.43	S5 (39.35)，S24 (98.2%)	6.07	S5	podophyllotoxin
7	33.6[a]	S5 (12.6%), S6 (73.3%), S7 (37.3)**, S8 (19.9%)**, S9 (16.3%)**, S10 (13.9%)**, S11 (15.55%),	11.63**	S6	kaemferol
8	33.7[a]	S7 (37.3)**, S8 (19.9%)**, S9 (16.3%)**, 10 (13.9%)**	11.63**	S7	podophyllotoxone
9	37.95	S8 (8.7%), S9 (19.6%)	3.37	S9	podoverine A
10	39.33	S8 (8.1%), S9 (18.2%)	3.66	S9	podoverine F
11	40.81	S10 (13.7%), S11 (5.6%)	0.95	S10	podoverine G
12	42.02	S11 (13.4%), S12 (50.9%), S13 (82.5%)	6.05	S13	podoverine D
13	44.63	S17 (55.3%), S18 (40%), S19 (8.5%)	3.89	S17	podoverine E
14	46.19	S19 (18.4%), S20 (64%)	1.39	S20	podoverine H
15	47.63	S21 (32.1%)	0.89	S21	podoverine I

[a] Due to minor difference of retention time, their relative contents was not accurately detected by DAD.

* The value is a possible sum of two overlapped compounds 2 and 3; and ** for the compounds 7 and 8.

3.3. NMR analyses and hierarchical clustering analysis of ^{13}C shifts

NMR is a powerful profiling tool for quantitative and qualitative analysis of natural products, in which ^{1}H and ^{13}C NMR spectrum are most frequently used. ^{1}H NMR is effective in the analysis of hydrogen atoms in organic compounds. However, in most cases there are many overlapping signals in the analysis of complex mixtures by ^{1}H NMR because of the limit of spectral width (0-15 ppm). But the situation is greatly improved in ^{13}C NMR, which has a much wider broadband (commonly from 0 ppm to 250 ppm), one single ^{13}C signal correspondence with one specific C atom position of the molecule.[36] Therefore, the scattered ^{13}C signals of the fractions containing different components can be clustered into different clusters that each clusters only contains the ^{13}C signals of one compound by use of the NMR-based pattern recognition technique.[37]

^{13}C NMR data of all fractions (Fig. Supp.1) were first processed as described in section 2.5, and a two-dimensional table (Table Supp.1) containing 24 columns (samples) and 146 rows (^{13}C signals) was obtained. The table was then subjected to the hierarchical clustering analysis using the software GENE-E. As shown in Fig. 3, four major blocks, including two podophyllotoxins blocks (i and iii) and two flavonoids blocks (ii and iv), were highlighted. These blocks contained several clusters *A* to *K*, which were assigned to molecular structures with the help of an on-line ^{13}C NMR database and the data of the metabolites of *D. versipellis* reported in the literatures.

3.4. Identification of the recognized ^{13}C NMR clusters

As is well known, the primary components of *D. versipellis* are podophyllotoxins and flavonoids. By analyzing the chemical shifts of the 2D heat map alone and combining the ^{13}C NMR characteristics of the structures of podophyllotoxins and flavonoids, four blocks (i-iv) were obtained. Red colour area with almost the same chemical shifts was viewed as one cluster. Considering the probability that the chemical shifts of one compound may not cluster together, clusters in different blocks but with the same kind of compounds (i and iii, ii and iv) were regarded as one cluster on the premise that the clusters were ruled out the possibility to represent other compounds. At a result, clusters A to K were obtained.

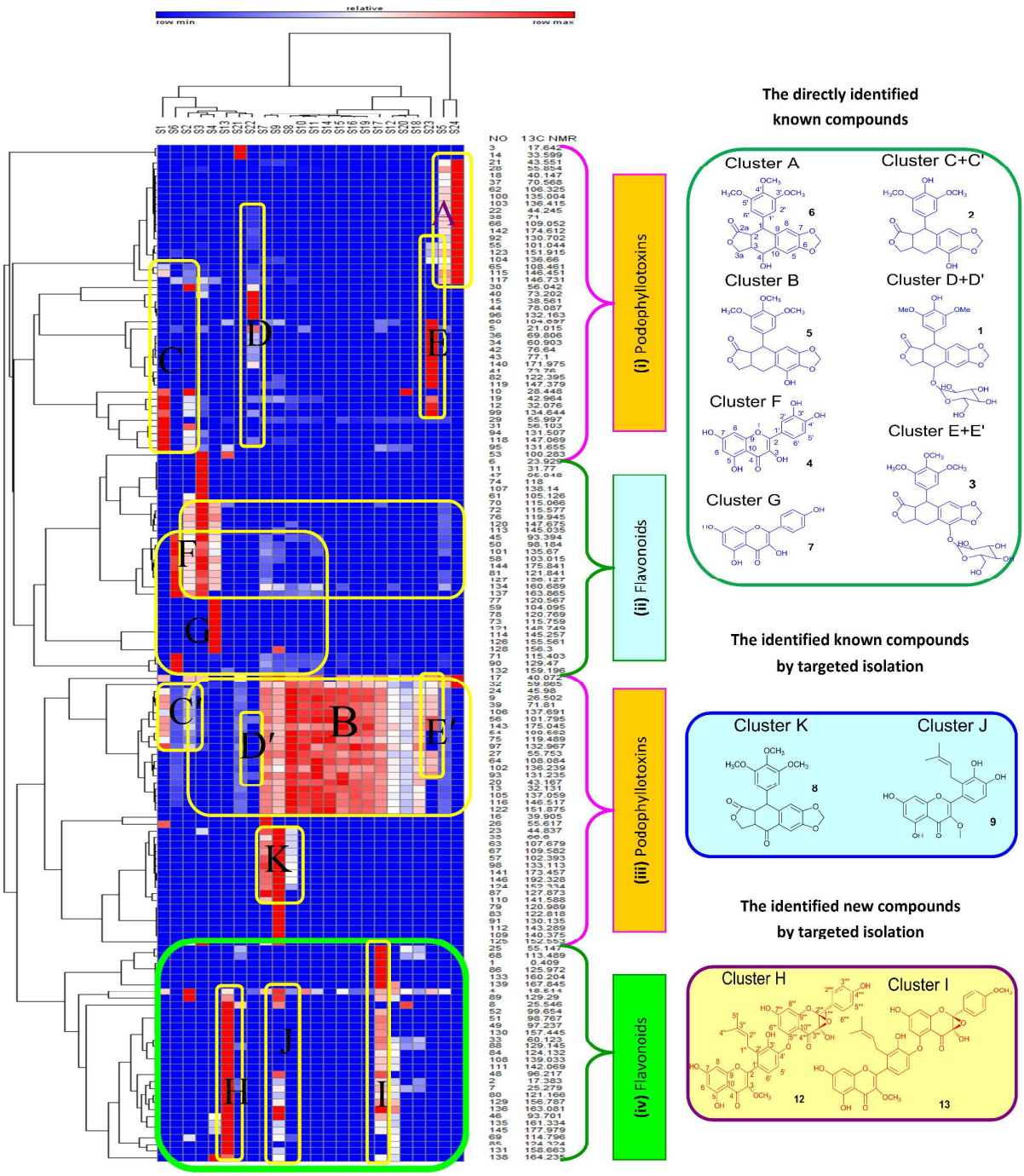


Fig. 3. ^{13}C NMR map and the ^{13}C shift clusters by hierarchical clustering analysis of the metabolites of RPLC fractions. The minimum intensity threshold of ^{13}C NMR signals of fractions was set at 0.005. The ^{13}C NMR chemical shifts were first clustered by hierarchical clustering analysis using a soft of GENE-E, then the clustered ^{13}C signals were submitted to an on-line structure search engine of MICRONMR database for structural determination. 7 prominent components (marked in blue) were directly recognized and four components were identified by following targeted isolation. Two flavone dimers (marked in red) were new natural products.

At first, we checked the cluster of the pure model compound of podophyllotoxin crystallized from the fraction S5. As shown in Fig. 3, it was well defined into cluster A of the block (i) podophyllotoxin together with the same component of the fraction S5. When the NMR signals were imported into the on-line structure search engine of MICRONMR database, podophyllotoxin was the optimal matching result with the similarity of 94.5%.[49] And we noticed that the signal of a carbon atom δ 59.87 ppm, (Fig.3, Table Supp. 2) was not clustered in cluster A (the reason will be discussed below). Thus the structure of the cluster A was first defined as podophyllotoxin.

Then the data of cluster B of the block (iii) podophyllotoxins, with more than half fractions containing it, was imported in the MICRONMR database, and it completely matched *β*-peltatin (similarity of 100%, Table S2). So the structure of the cluster B was defined as *β*-peltatin.[50]

In the same way, cluster G (fraction S5-S11) and F (fraction S2-S11) of the block (ii) flavonoids matched the compounds kaemferol and quercetin with the similarity of 92.3%, 93.3%,[50, 51] respectively (Table Supp. 3). Thus, the structure of the cluster G was defined as kaemferol and F was defined as quercetin.

It should be noted that the ^{13}C signals of some compounds might be split into two or more sub-clusters due to the structural similarity which resulted in overlapped ^{13}C NMR shifts of different compounds. For examples, as shown in Fig. 3, the cluster C of the block (i) podophyllotoxins and cluster C′ of the block (iii) podophyllotoxins (including fraction S1 and S2) seemed to correspond to one compound of podophyllotoxins. All ^{13}C signals clustered in the cluster C and C′ were found to well match the data of *α*-peltatin[52] with similarity of 85%. Thus the structure of the clusters C and C′ was defined as *α*-peltatin.

Similarly, the clusters D of the block (i) podophyllotoxins and cluster D′ of the block (iii) podophyllotoxins (fraction S22) were found to be two split clusters. Its possible structure was 4′-demethylpodophyllotoxin glucoside **(**similarity of 84.6%[53]) by on-line screening in MICRONMR database.

In addition, although the fraction S23 had high purity (91.8% in peak area ratio), its ^{13}C signals were split into two sub-clusters E and E′ in the block (i) and (iii). After online ^{13}C screening, these ^{13}C signals found to be well match the data of *β*-peltatin glucoside[52]. Thus the clusters E and E′ were defined as *β*-peltatin glucoside.

However, when the ^{13}C signals of one compound were split into more than one clusters, it made the structural assignment difficult. For example, some signals of the cluster K were split into other clusters (Fig. 3, Table S2), which led to that its structure was difficult to be defined. In addition, for the block (iv), although the ^{13}C signals of clusters H, I, and J indicated that they possibly had flavone subunit, the clustered ^{13}C signals could not target any structures by on-line NMR database screening, which suggested that they were possibly new natural products. Therefore, these clusters were further identified by following targeted isolation.

3.5 Validation of recognized metabolites by targeted purification

As described above, the pure podophyllotoxin as model sample (S24) to validate the method was crystallized from fraction S5. The purified podophyllotxin showed negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 413 and positive ESI-TOF-MS ion $[M+Na]^+$ at *m/z* 437.1210, implying that its molecular weight was 414 and molecular formula was $C_{22}H_{22}O_8$. On the other hand, the NMR data (Table Supp. 2) are closely in agreement with previous reported data.[49]. Therefore, the peak 6 was identified as podophyllotxin for cluster A.

As shown in Fig. 3, the ^{13}C signals clustered in the cluster B were found in the fractions of S4-S12 and S14-S21 (Table 1). Among these fractions, the fractions of S8, S10, S11, S14, S15, S16 and S19 had stronger intensity of ^{13}C signals (deeper red colour in the map), which implied a higher content of the compound. Together with HPLC analysis (Fig. 2), peak 5 was the most possible compound of the cluster B. Thus, peak 5 was used as a target for purification. According to the HPLC content (Table 1) and ^{13}C NMR intensities in cluster B (Fig. 3), S14 was the most suitable fraction for the preparation of peak 5 because of the deepest red colour and the least interference signals. So S14 was selected as sample for further targeted purification. The purified peak 5 gave negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 413 and positive ESI-TOF-MS ion $[M+Na]^+$ at *m/z* 437.1206, implying that its molecular weight was 414 and molecular formula was $C_{22}H_{22}O_8$. And its NMR data agreed with the reported data[54] and the signals in cluster B (Table S2). Therefore, the cluster B was well defined as *β*-peltatin (**5**).

Above NMR heat map analyses (Fig. 3) suggested that the ^{13}C signals of clusters C+C′ possibly corresponded to *α*-peltatin. In addition, the heat-map (Fig.3) indicated that these signals contained in the fractions of S1 and S2, but only S1 had little other signals. HPLC analyses (Fig. 2) indicated the two fractions contained a common peak of peak 2. Therefore, S1 was used as the sample for targeted purification of peak 2. The purified peak 2 gave negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 399 and positive ESI-TOF-MS ion $[M+Na]^+$ at *m/z* 423.1035, implying that its molecular weight was 400 and molecular formula was $C_{21}H_{20}O_8$.The NMR data of the purified product (Table Supp. 2) confirmed that the clusters C+ C′ were identified as *α*-peltatin (**2**).[55]

Similarly, the fraction S3 was selected as the sample for the targeted separation of peak 4, which corresponded to cluster F. The purified peak 4 gave negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 301 and ESI-TOF-MS ion $[M-H]^-$ at *m/z* 301.0346, implying that its molecular weight was 302 and molecular formula was $C_{15}H_{10}O_7$. And its NMR data (Table Supp. 3) fit the data of quercetin in previous report.[50] Thus cluster F was identified as quercetin (**4**).

HPLC analyses (Fig. 2) showed that the purity of fraction S22 was not satisfactory and the main peak 1 had purity of about 58%. However, there were little ^{13}C signals out of the clusters D+D′, implying that the actual contents of some peaks in the HPLC profile were very little. Therefore, the fraction S22 was selected as the sample and peak 1 was isolated. As a result, the ^{13}C signals of prepared peak 1 were closely in agreement with the shifts of clusters D+D′. The purified peak 1 gave negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 561 and negative ESI-TOF-MS ion $[M-H]^-$ at *m/z* 561.1631, implying that its molecular weight was 562 and molecular formula was $C_{27}H_{30}O_{13}$.Thus the combination of clusters D+D′ was well identified as 4′-demethypodophyllotoxin glucoside (**1**).[53]

Fig. 3 showed that almost all of the ^{13}C signals of the fraction S23 were clustered in clusters E+E′. Furthermore, there was only one peak 3 in the HPLC profile of the fraction (Fig. 2). Thus the peak 3 was a possible target for further purification from the fraction S23. The purified peak 3 gave the structural data matching with the reported data of *β*-peltatin glucoside.[52] The purified peak 3 gave

negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 575 and positive ESI-TOF-MS ion $[M+Na]^+$ at *m/z* 599.1735, implying that its molecular weight was 576 and molecular formula was $C_{28}H_{32}O_{13}$.Thus the clusters E+E′ were well identified as *β*-peltatin glucoside (**3**).

The deepest red colours of ^{13}C signals in the cluster G were found in the fraction S6. HPLC analysis indicated that the fraction S6 contained at least three peaks 4, 5, and 7 and peak 7 had more than 70% area ratio. As described above, peaks 4 and 5 had been identified as quercetin and *β*-peltatin, respectively. Thus, peak 7 was selected as the target for further preparation. As a result, the purified peak 7 showed the same structural data (Table Supp. 3) with the reported data of kaemferol.[51]. The purified peak 7 gave negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 285 and negative ESI-TOF-MS ion $[M-H]^-$ at *m/z* 285.0410, implying that its molecular weight was 286 and molecular formula was $C_{15}H_{10}O_6$.Therefore, the cluster G was identified as kaemferol (**7**).

3.6 Targeted isolation and identification of the un-recognized ^{13}C NMR clusters for the discovery of new natural products

As described above, the ^{13}C signals of cluster K and one block including three clusters H, I and J at least could not be recognized by on-line ^{13}C NMR database. Therefore, these clusters were further purified and identified from the pre-fractionized fractions.

As shown in Fig. 3, the ^{13}C signals of three fraction samples of S7, S8 and S9 were clustered in the cluster K. HPLC analyses showed that the peaks of 4, 5, 7 and 8 were the major components. Because the peaks of 4, 5 and 7 had been identified above, the peak 8 was a possible target for further isolation. The NMR map analysis (Fig. 3) indicated that the fractions S7 and S9 had deeper red colour, implying higher content of the target. Thus we selected fraction S7 as a sample. As a result, the purified peak 8 showed negative ESI-MS molecular ion $[M-H]^-$ at *m/z* 411 and positive ESI-MS molecular ion $[M-H]^+$ at *m/z* 413, implying that its molecular weight was 412. The NMR data was closely agreement with the reported data of podophyllotoxone.[53] Therefore, the cluster K was identified as podophyllotoxone (**8**).

Fig. 3 showed that the cluster H seemed an integral ^{13}C NMR shifts collection but its possible structure could not be found from the on-line MICRONMR database and published references. In addition, the signals clustered in the cluster H were major ^{13}C signals of fraction S13 (Fig. 3). Thus the cluster H was defined to correspond to the peak 12, which was further purified from the fraction S13.

Besides the cluster H, there were two clusters I and J contained in block (iv), which were some split. Their structures also could not be defined directly from the database. Thus the cluster I (peak 13) was prepared from the fraction S17 while cluster J (peak 9 or 10) was purified from the fraction S9.

Due to the coexistence in the same block (iv), the clusters H, I and J had some similar ^{13}C signals, implying that they had the same groups in part of their structures. The NMR data (Table Supp. 3) of purified components indicated that there were about half of ^{13}C signals of H and I are the same with the ones of J, which suggested that J was a monomer of H and I. Therefore the structure of J was first addressed. On-line ^{13}C data screening indicated that the compound was a known compound: podoverine A (Fig. Supp. 2 and Table Supp. 4), which was isolated from *Podophyllum versipelle* Hance[55] and had been found to be a novel microtubule destabilizing agent recently.[56].

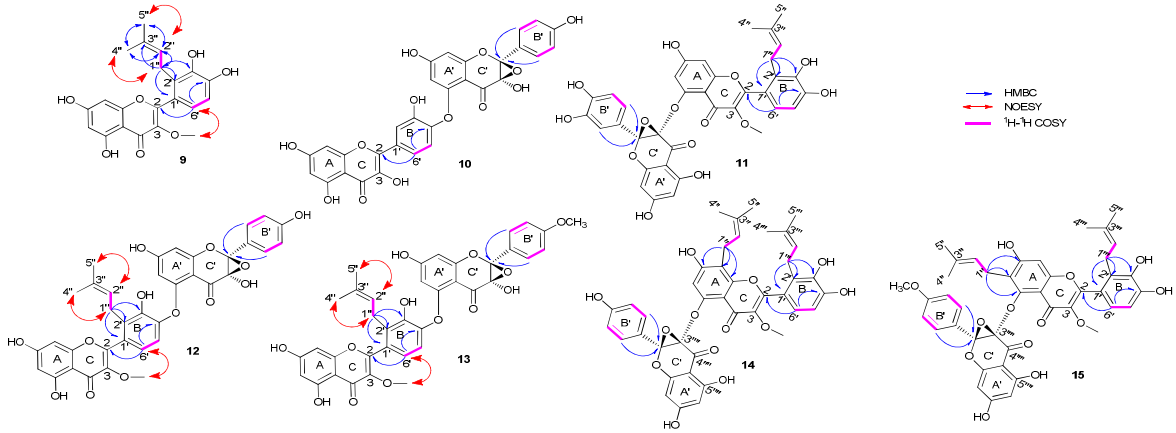


Fig. 4. The chemical structures of compounds **9**-**15** and their Key ^1H-^1H COSY, HMBC and NOESY signals.

The high resolution ESI-TOF-MS analysis of **12** indicated that its molecular formula was $C_{36}H_{28}O_{13}$. The positive ESI-MS^2 spectrum showed the compound **12** was easily broken into two monomer with positive ions at *m/z* 383 (100%) and *m/z* 287 (22%). Similar monomers were observed at *m/z* 381 (13%) and 283 (100%) in the negative ESI-MS spectrum (Fig. Supp. 3A-E). 1D and 2D NMR data (Table Supp. 5) also suggested that it was a flavone dimmer and the two subunits were identified as podoverine A and 3-OH-(2,3-epoxy)flavone. The structure of **12** was established as Fig. 4. Because it was a dimer of podoverine A and an analogue of podoverine B and C were isolated from *Podophyllum versipelle* Hance,[55] we named the compound **12** as podoverine D.

^1H and ^{13}C NMR data, and ESI-MS analysis (Fig. Supp. 4 and Table Supp. 6) showed that there were minor differences between compounds **12** and **13** except for an OCH_3 at C4′′′′ position (δ_C 55.15, δ_H 3.75). The key coupling correlations were illustrated in the Fig. 4. All assignments of H and C atoms were confirmed by 1D and

2D NMR data (Table. Supp. 6). Therefore, the structure of **13** was established and named as podoverine E.

3.7 Targeting isolation and identification of minor new flavone dimers enriched by RPLC fractionation

There were two peaks which could match the data of cluster J, so they both purified from S9. And the results implied that peak 9 corresponded to cluster J (the signals of peak 10 in S9 were too weak to emerge in the 2D map). Further isolation and identification proved that compound **10** corresponding peak 10 was also a new flavone dimer.

When the minimum intensity threshold of ^{13}C NMR signals of fractions was set at 0.005 (Fig. 3), two new flavone dimers were found and isolated. But ^{13}C NMR spectra of fractions (Fig. Supp.1) and HPLC spectra (Fig. 1 and 2) indicated that there were still some minor un-identified components in the fraction samples such as S20 and S21. Although the contents of these components in the crude sample were very low (Fig. 1), their contents in the RPLC fractions (Fig. 2) seemed a significant increase. However, in the level of 0.005, the most of their ^{13}C NMR signals were not detected. For example, the area percentage of peak 14 was as high as 64% of peak area in fraction S20, peak 15 was 32% of peak area, but the most of their ^{13}C NMR signals did not appear in Fig. 3. Therefore, the lower level of 0.002 (higher than 3 S/N) was used.

As shown in Fig. Supp. 5 and 6, there were more ^{13}C signals to be clustered. Similar to Fig. 3, the ^{13}C NMR map showed same ^{13}C shift blocks including two podophyllotoxins and two flavonoids, and several clusters. Interestingly, the cluster K were clearly grouped and the block for flavones dimers (Fig. 5) showed more clusters, including cluster L, M and N besides cluster H, I and J, suggesting that they might be also new flavones dimers. So peak 11 in S10 (cluster L), peak 14 in S20 (cluster M) and peak 15 in S21 (cluster N) were further isolated.

ESI-MS spectra of compound **10** showed that it had prominent ions of $[M-H]^-$ at *m/z* 585 (82%), $[2M-H]^-$ at *m/z* 1171 (100%) and $[M+H]^+$ at *m/z* 587 (100%) (Fig. Supp.7A-D), suggesting that its molecular weight was 586. The high resolution ESI-TOF-MS analysis indicated that its molecular formula was $C_{30}H_{18}O_{13}$ (Fig. Supp. 7E). Its 1D and 2D NMR data were similar with those of compounds **12** and **13**, but there was no prenyl group signals. The structure of **10** was illustrated in the Fig. 4, which was fully supported by 1D and 2D NMR data (Table Supp. 7). Similarly, it was named as podoverine F.

ESI-MS spectra of compound **11** showed that it had prominent ions of $[M-H]^-$ at *m/z* 683 (100%), $[2M-H]^-$ at *m/z* 1367 (34%) and $[M+H]^+$ at *m/z* 685 (75%), $[2M+Na]^+$ at *m/z* 1391 (100%), suggesting that its molecular weight was 684. The high resolution ESI-TOF-MS analysis indicated that its molecular formula was $C_{36}H_{28}O_{14}$(Fig. Supp.8A-E). The structure of **11** was established as Fig. 4. Its 1D and 2D NMR data were summarized in Table Supp. 8. And it was named as podoverine G.

ESI-MS spectra of compound **14** showed that it had prominent ions of $[M-H]^-$ at *m/z* 735 (100%), $[2M-H]^-$ at *m/z* 1471 (88%) and $[M+H]^+$ at *m/z* 737 (100%), $[2M+Na]^+$ at *m/z* 1391 (40%) (Fig. Supp.9A-D), suggesting that its molecular weight was 736. The high resolution ESI-TOF-MS analysis indicated that its molecular formula was $C_{41}H_{36}O_{13}$(Fig. Supp.9E). ESI-MS/MS analysis indicated it also contained two moieties. And the structure of **14** was established as Fig. 4. Its 1D and 2D NMR data were summarized in Table Supp. 9 and it was named as podoverine H.

ESI-MS spectra of compound **15** showed that it had prominent ions of $[M-H]^-$ at *m/z* 749 (100%), $[2M-H]^-$ at *m/z* 1499 (14%) and $[M+H]^+$ at *m/z* 751 (100%), suggesting that its molecular weight was 750. The high resolution ESI-TOF-MS analysis indicated that its molecular formula was $C_{42}H_{38}O_{13}$ (Fig. Supp.10A-E). ESI-MS/MS analysis indicated it also contained two moieties. The structure of **15** was established as Fig. 4. Its NMR spectra were summarized in Table Supp.10 . And it was named as podoverine I.

The absolute configuration of new compounds 10-15 were determined by CD measurements (Fig. Supp.11) and illustrated in Fig. 4.

3.8 Evaluation of anti-cancer efficiency in vitro

At last, the cytostatic activities of purified compounds were evaluated against several cancer cells *in vitro,* including human breast cancer Bcap37, human hepatoma HepG2, doxorubicin-resistant human hepatoma R-HepG2, mouse melanoma B16 and brain glioma GL261 cells. All the compounds showed efficient activities. Their IC_{50} was only several μg/mL or even several ng/mL (Table 2).

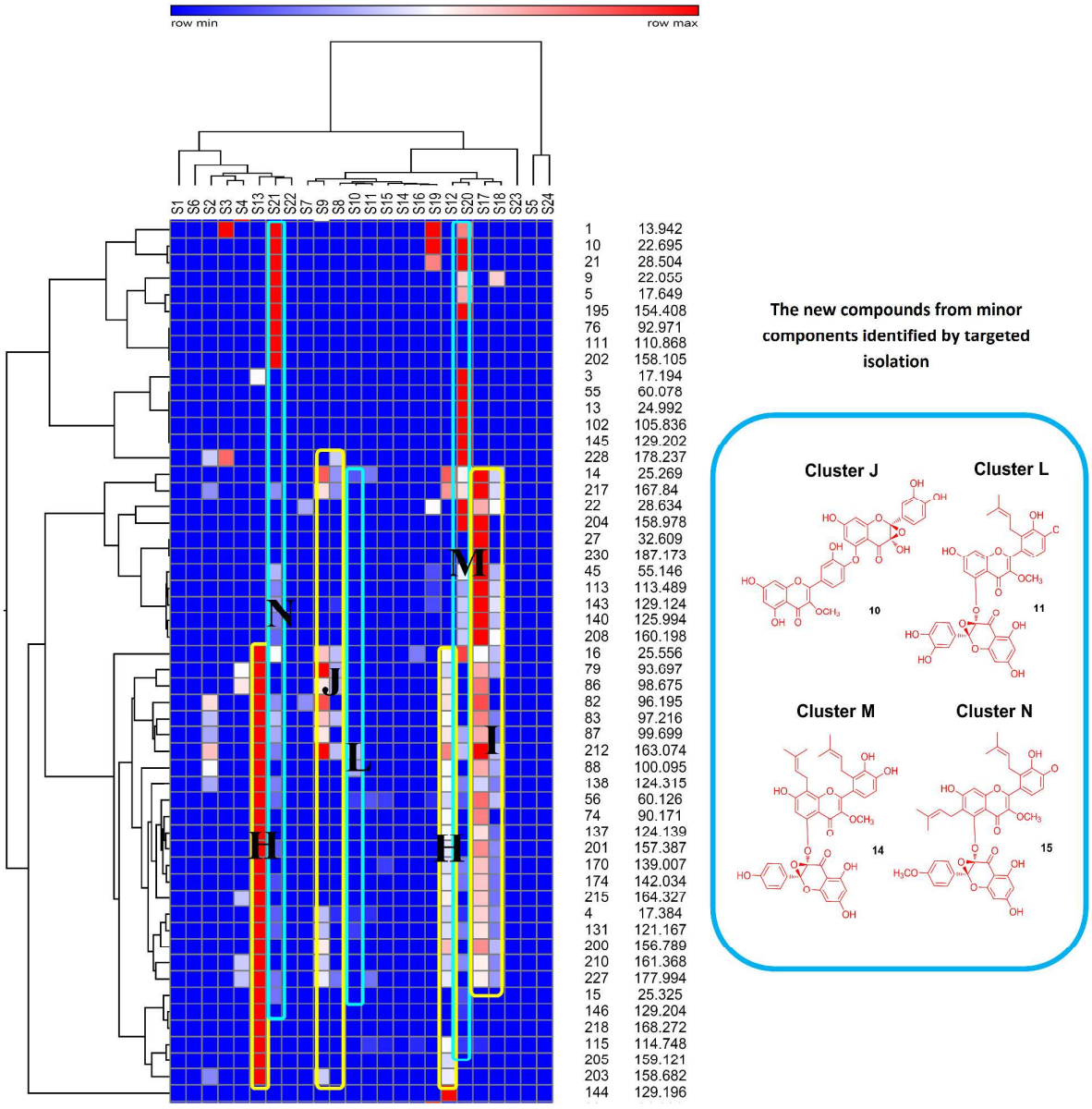


Fig. 5. Expanded view of flavones dimers block in the ^{13}C NMR map (Fig. Supp. 5) on the level of 0.002.

Journal Name

ARTICLE

4. Conclusions

In conclusion, this work provided a successful strategy by combination of RPLC fractionation-^{13}C NMR pattern recognition to rapidly recognize the known components and discover new natural products from natural products. Generally speaking, as shown in Fig. 6, this strategy contained four steps, including (1) preparation of crude sample, (2) RPLC or LC fractionation of crude sample, (3) ^{13}C NMR measurement of each fraction and following ^{13}C NMR map-based hierarchical clustering analyses, (4) identification of known components and discovery of new compounds.

Table 2. Cytotoxicity of identified compounds on different cancer cell lines with 48 h drug exposure.

No.	Compounds (concentration)	Cytotoxicity (IC_{50})				
		B16	HepG2	R-HepG2	GL261	BCap37
1	4'-demethylpodophyllotoxin glucoside (μg/mL)	13.48	14.19	27.11	31.02	8.71
2	α-peltatin (ng/mL)	25.28	8.82	605.30	41.76	1532.00
3	β-peltatin glucoside (μg/mL)	14.63	0.97	8.40	8.77	3.65
4	quercetin (μg/mL)	6.74	0.29	0.58	14.67	0.16
5	β-peltatin (ng/mL)	5.55	2.62	13.80	10.53	2.87
6	podophyllotoxin (ng/mL)	474.30	8.00	20.52	13.61	< 5
7	kaemferol (μg/mL)	5.01	3.44	10.76	5.47	3.93
8	podophyllotoxone (μg/mL)	0.57	< 0.1	0.13	0.35	< 0.1
9	podoverine A (μg/mL)	7.20	0.13	4.40	19.92	< 0.1
10	podoverine F (μg/mL)	5.34	0.24	3.42	9.50	< 0.5
11	podoverine G (μg/mL)	3.93	2.30	4.23	6.94	0.95
12	podoverine D (μg/mL)	1.79	< 0.5	0.99	3.88	< 0.5
13	podoverine E (μg/mL)	4.25	4.54	8.79	5.01	4.82
14	podoverine H (μg/mL)	6.77	0.13	2.94	3.69	1.72
15	podoverine I (μg/mL)	10.70	6.29	9.70	9.39	7.76

As an example, we identified seven known prominent components of *D. versipellis* directly from 7 single clusters or cluster-combinations of ^{13}C NMR signals of the RPLC fractions through ^{13}C NMR map-based hierarchical clustering analyses. Furthermore, from unrecognized ^{13}C NMR clusters, we isolated and identified six new flavones dimers along with two known compounds. Interestingly, all compounds showed potent anti-cancer activities.

Thus the developed strategy of RPLC fractionation- ^{13}C NMR pattern recognition proved to be a very practical method to identify the known and new metabolites from complex natural product extracts. Because RPLC or LC or other column chromatography is a very popular separation technique, the LC fractionation-NMR pattern recognition strategy can be explored as a standard lab protocol for natural products development. In addition, it will also be useful in complex metabolomic analyses, especially in the fields of identification and characterization of the fingerprints and search for new components from complex Traditional Chinese Medicines or other natural resources. If the larger RPLC or other chromatographic columns with higher resolution are used to prepare the more resolved the fractions from the larger-scale sample mass, it is very possible to discover the minor or trace amounts new natural products.

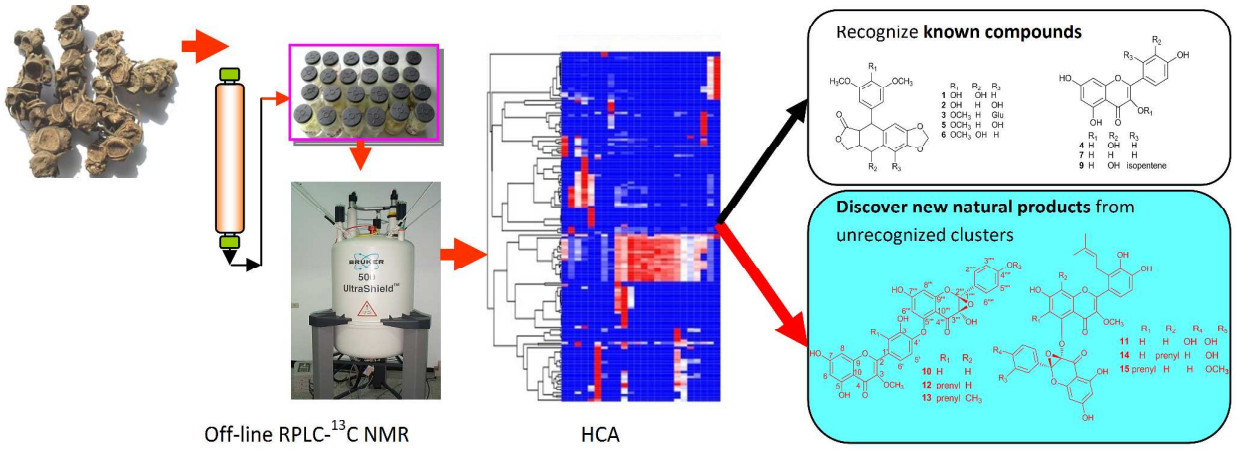


Fig. 6. A general protocol of the combination strategy of the RPLC fractionation-^{13}C NMR pattern recognition for the recognition known components and discover the new natural products.

Acknowledgements

We would like to thanks Prof. Shuqun Zhang for valuable suggestions to improve the manuscript.

This work was supported in part by National Natural Science Foundation of China (grant no.: 21272209, 20972136).

References

1 A. Ganesan, *Current Opinion in Biotechnology*, 2004, **15**, 584-590.
2 J.-Y. Ortholand and A. Ganesan, *Current opinion in chemical biology*, 2004, **8**, 271-280.
3 J. W. H. Li and J. C. Vederas, *Science*, 2009, **325**, 161-165.
4 G. Appendino, A. Minassi and O. Taglialatela-Scafati, *Natural Product Reports*, 2014, **31**, 880-904.
5 L. Crombie, *Pesticide Science*, 1999, **55**, 761-774.
6 A. T. Bull and J. E. M. Stach, *Trends in Microbiology*, 2007, **15**, 491-499.
7 A. Gosslau, S. M. Li, C. T. Ho, K. Y. Chen and N. E. Rawson, *Molecular Nutrition & Food Research*, 2011, **55**, 74-82.
8 D. J. Newman and G. M. Cragg, *Journal of Natural Products*, 2012, **75**, 311-335.
9 M. A. Rudek, C. H. Chau, W. D. Figg and H. L. McLeod, *New York, NY : Springer New York*, 2014, 39-68.
10 J. W. Allwood and R. Goodacre, *Phytochemical Analysis*, 2010, **21**, 33-47.
11 A. Bouslimani, L. M. Sanchez, N. Garg and P. C. Dorrestein, *Natural Product Reports*, 2014, **31**, 718-729.
12 D. Krug and R. Muller, *Natural Product Reports*, 2014, **31**, 768-783.
13 R. C. Breton and W. F. Reynolds, *Natural Product Reports*, 2013, **30**, 501-524.
14 F. Bucar, A. Wube and M. Schmid, *Natural Product Reports*, 2013, **30**, 525-545.
15 M. Zerikly and G. L. Challis, *Chembiochem*, 2009, **10**, 625-633.
16 A. Harvey, *Drug Discovery Today*, 2000, **5**, 294-300.
17 O. Corcoran and M. Spraul, *Drug Discovery Today*, 2003, **8**, 624-631.
18 J. C. Lindon, *Drug Discovery Today*, 2003, **8**, 1021-1022.
19 M. Godejohann, *Planta Medica*, 2010, **76**, 1175-1175.
20 R. L. Last, A. D. Jones and Y. Shachar-Hill, *Nature Reviews Molecular Cell Biology*, 2007, **8**, 167-174.
21 H. K. Kim, Y. H. Choi and R. Verpoorte, *Nature Protocols*, 2010, **5**, 536-549.
22 G. F. Pauli, B. U. Jaki and D. C. Lankin, *Journal of Natural Products*, 2005, **68**, 133-149.
23 D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Research*, 2013, **41**, D801-D807.
24 T. Jewison, C. Knox, V. Neveu, Y. Djoumbou, A. C. Guo, J. Lee, P. Liu, R. Mandal, R. Krishnamurthy, I. Sinelnikov, M. Wilson and D. S. Wishart, *Nucleic Acids Research*, 2012, **40**, D815-D820.
25 L. Shanghai Micronmr Infor Technology Co., ,http://www.nmrdata.com:90/.
26 http://c13.usal.es/c13/usuario/views/inicio.jsp?lang=en&country=EN.
27 J. K. Nicholson, J. C. Lindon and E. Holmes, *Xenobiotica*, 1999, **29**, 1181-1189.
28 W. Weckwerth and K. Morgenthal, *Drug Discovery Today*, 2005, **10**, 1551-1558.
29 S. L. Robinette, R. Bruschweiler, F. C. Schroeder and A. S. Edison, *Accounts of Chemical Research*, 2012, **45**, 288-297.
30 J. K. Nicholson, J. Connelly, J. C. Lindon and E. Holmes, *Nature Reviews Drug Discovery*, 2002, **1**, 153-161.
31 J. L. Griffin, *Current opinion in chemical biology*, 2003, **7**, 648-654.
32 M. Aursand, I. B. Standal and D. E. Axelson, *Journal of Agricultural and Food Chemistry*, 2007, **55**, 38-47.
33 M. Aursand, I. B. Standal, A. Prael, L. McEvoy, J. Irvine and D. E. Axelson, *Journal of Agricultural and Food Chemistry*, 2009, **57**, 3444-3451.
34 A. M. Sarotti, *Organic & Biomolecular Chemistry*, 2013, **11**, 4847-4859.
35 H. C. Keun, O. Beckonert, J. L. Griffin, C. Richter, D. Moskau, J. C. Lindon and J. K. Nicholson, *Analytical Chemistry*, 2002, **74**, 4588-4593.
36 J. Hubert, J. M. Nuzillard, S. Purson, M. Hamzaoui, N. Borie, R. Reynaud and J. H. Renault, *Analytical Chemistry*, 2014, **86**, 2955-2962.
37 S. K. Oettl, J. Hubert, J. M. Nuzillard, H. Stuppner, J. H. Renault and J. M. Rollinger, *Analytica Chimica Acta*, 2014, **846**, 60-67.
38 J. V. Marques, K.-W. Kim, C. Lee, M. A. Costa, G. D. May, J. A. Crow, L. B. Davin and N. G. Lewis, *The Journal of biological chemistry*, 2013, **288**, 466-479.
39 Y.-Q. Liu, J. Tian, K. Qian, X.-B. Zhao, S. L. Morris-Natschke, L. Yang, X. Nan, X. Tian and K.-H. Lee, *Medicinal research reviews*, 2015, **35**, 1-62.
40 X. Xu, X. Gao and L. Jin, *Cell Div.*, 2011, **6**, 14.
41 Y. Sun, Z. Li, H. Chen, W. Zhou and H. Hua, *Chinese Trad. Herb. Drugs*, 2012, **43**, 1626-1634.
42 Y. Q. Liu, L. Yang and X. Tian, *Curr. Bioactive Compounds*, 2007, **3**, 37-66.
43 T. J. Schmidt, S. Hemmati, E. Fuss and A. W. Alfermann, *Phytochemical analysis : PCA*, 2006, **17**, 299-311.
44 Y. Zhou, S. Y. Jiang, L. S. Ding, S. W. Cheng, H. X. Xu, P. P. H. But and P. C. Shaw, *Chromatographia*, 2008, **68**, 781-789.
45 S. K. Wong, S. K. Tsui, S. Y. Kwan, X. L. Su and R. C. Lin, *Journal of mass spectrometry : JMS*, 2000, **35**, 1246-1251.
46 Z. Yang, X. M. Liu, K. W. Wang, X. J. Cao and S. H. Wu, *Journal of Separation Science*, 2013, **36**, 1022-1028.

47 H. Zhou, J. L. Liang, D. Lv, Y. F. Hu, Y. Zhu, J. P. Si and S. H. Wu, *Food Chemistry*, 2013, **138**, 2390-2398.
48 A. L. Harvey, R. Edrada-Ebel and R. J. Quinn, *Nature Reviews Drug Discovery*, 2015, **14**, 111-129.
49 Y. Ma, *J. Anhui Agri. Sci.*, 2011 **39**, 19781-19782
50 X. M. Ma, Y. Liu and Y. P. Shi, *Chemistry & Biodiversity*, 2007, **4**, 2172-2181.
51 K. Masanori, F. Masamichi, Y. Kunitoshi, N. Shinsaku and Y. Kazuo, *Chemical & pharmaceutical bulletin*, 1978, **26**, 3594-3596.
52 M. A. Rashid, K. R. Gustafson, J. H. Cardellina and M. R. Boyd, *Natural Product Letters*, 2000, **14**, 285-292.
53 F. Jiang, *Chinese Trad. Herb. Drugs*, 2011, **42**, 634-639.
54 F. A. Ramos, Y. Takaishi, M. Shirotori, Y. Kawaguchi, K. Tsuchiya, H. Shibata, T. Higuti, T. Tadokoro and M. Takeuchi, *Journal of Agricultural and Food Chemistry*, 2006, **54**, 3551-3557.
55 H. Arens, B. Ulbrich, H. Fischer, M. J. Parnham and A. Romer, *Planta Medica*, 1986, 468-473.
56 T. T. N. Tran, C. Gerding-Reimers, B. Scholermann, B. Stanitzki, T. Henkel, H. Waldmann and S. Ziegler, *Bioorganic & Medicinal Chemistry*, 2014, **22**, 5110-5116.