Anal. Chem. **2008,** *80,* 8187–8194

Global Detection and Identification of Nontarget Components from Herbal Preparations by Liquid Chromatography Hybrid Ion Trap Time-of-Flight Mass Spectrometry and a Strategy

Haiping Hao,[†] Nan Cui,[†] Guangji Wang,*[,†] Binren Xiang,[‡] Yan Liang,[†] Xiangyang Xu,[§] Hui Zhang,[§] Jun Yang,[§] Chaonan Zheng,[†] Liang Wu,[†] Ping Gong,[†] and Wei Wang[†]

Key Laboratory of Drug Metabolism and Pharmacokinetics, Center for Instrumental Analysis, China Pharmaceutical University, Nanjing 210009, P.R. China, and Jinling Pharmaceutical Co., Ltd, Nanjing 210009, P.R. China

Although the current literature has recorded many reports of identifying components from herbal preparations, all of them were largely limited to target components. This paper provides a novel and generally applicable approach to identifying nontarget components from herbal preparations, based on the use of liquid chromatography ion trap time-of-flight mass spectrometry (LC/MS-IT-TOF). A simple program was originally developed for searching the common diagnostic ions from all experimentally generated ions. The components sharing the exact same ions (mass error < 5 mDa) were classified into a family. All families were then connected into a coherent network by the bridging components that are present in two or more families. With the benefit from such a network, it is feasible to sequentially characterize the structures of all diagnostic ions once a single component has been de novo identified. The structures of the diagnostic ions could then be used as "a priori" information for selecting the exact candidates containing the substructures of the corresponding diagnostic ions from the primary database hits. This strategy enables a nearly 7-fold narrowing of the database hits and thus substantially enhances the analytical efficiency and sharpness. With the use of such an approach, 43 out of 53 components incorporated into the network have been successfully identified from the test herbal preparation. For the rest, components failed to be identified using this approach; a complementary approach to screening by sequential loss of specific chemical groups, proposed from the accurate mass differences between fragments, was established to narrow the database hits. All of the 87 peaks detected have been successfully identified by combining the use of both approaches except failed to differentiate some isomers. The presently developed approach and methodology would be useful for the identifications of complicated nontarget components from various complex mixtures such as herbal preparations, biological, and environmental samples.

As one of the oldest continuously practiced systems of traditional medicine in the world, herbal medicine has a history of several thousand years and their worldwide utilization has increased recently in both developing and developed countries.[1−4] The World Health Organization estimated that 65−80% of the world population used herbal medicines as the primary form of healthcare.[5]

It has been well acknowledged that for herbal medicines, the identification of components contained is of great significance to their quality control and to the disclosure of the secret underlying their effectiveness. Accordingly, qualitative and quantitative determinations of components contained in herbal medicines have now become a very hot issue. However, the rapid and reliable identification of chemical components contained in herbal preparations remains still a great challenge, despite recent advances in various analytical technologies. Although the previously reported methodologies could successfully identify up to dozens of components from herbal preparations,[6−15] most of them were limited to target components and depended largely on the use of reference compounds and/or the comparisons with the literature data. Considering that the reference compounds are always difficult to obtain and most components contained in herbal preparations are unknown (nontarget), the previously reported methods are apparently insufficient to the global detection and identification of the complicated components in herbal prepara-

* Corresponding author. Phone: +86-25-83271128. Fax: +86-25-85306750. E-mail: guangjiwang@hotmail.com.

[†] Key Laboratory of Drug Metabolism and Pharmacokinetics, China Pharmaceutical University.

[‡] Center for Instrumental Analysis, China Pharmaceutical University.

[§] Jinling Pharmaceutical Co.

(1) Lee, K. H. *Public Health Nutr.* **2000**, *3*, 515–522.
(2) Hu, Z.; Yang, X.; Ho, P. C.; Chan, S. Y.; Heng, P. W.; Chan, E.; Duan, W.; Koh, H. L.; Zhou, S. *Drugs* **2005**, *65*, 1239–1282.
(3) Bent, S.; Ko, R. *Am. J. Med.* **2004**, *116*, 478–485.
(4) Kaufman, D. W.; Kelly, J. P.; Rosenberg, L.; Anderson, T. E.; Mitchell, A. A. *J. Am. Med. Assoc.* **2002**, *287*, 337–344.
(5) World Health Organization. http://www.who.int/mediacentre/factsheets/fs134/en/.
(6) Wang, D. W.; Liu, Z. Q.; Guo, M. Q.; Liu, S. Y. *J. Mass Spectrom.* **2004**, *39*, 1356–1365.
(7) Drasar, P.; Moravcova, J. *J. Chromatogr., B* **2004**, *812*, 3–21.
(8) Cai, Z. W.; Lee, F. S. C.; Wang, X. R.; Yu, W. J. *J. Mass Spectrom.* **2002**, *37*, 1013–1024.
(9) Wang, P.; Liang, Y. Z.; Zhou, N.; Chen, B. M.; Yi, L. Z.; Yu, Y.; Yi, Z. B. *J. Mass Spectrom.* **2007**, *21*, 99–106.
(10) Gu, M.; Zhang, G. F.; Su, Z. G.; Yang, F. O. *J. Chromatogr., A* **2004**, *1041*, 239–243.
(11) Van Berkel, G. J.; Tomkins, B. A.; Kertesz, V. *Anal. Chem.* **2007**, *79*, 2778–2789.
(12) Chen, X. G.; Kong, L.; Su, X. Y.; Pan, C. S.; Ye, M. L.; Zou, H. F. *J. Chromatogr., A* **2005**, *1089*, 87–100.
(13) Zhang, Y.; Shi, Q.; Shi, P.; Zhang, W.; Cheng, Y. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 2328–2342.
(14) Zhu, Z. Y.; Zhang, H.; Zhao, L.; Dong, X.; Li, X.; Chai, Y. F.; Zhang, G. Q. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1855–1865.
(15) Qu, J.; Liang, Q. L.; Luo, G. A.; Wang, Y. M. *Anal. Chem.* **2004**, *76*, 2239–2247.

10.1021/ac801356s CCC: $40.75
Published on Web 09/17/2008

tions. Therefore, developing more powerful analytical tools and methods for the global characterizations of the chemical constituents in herbal preparations is of great concern.

Various hyphenated and hybrid mass spectrometers have now been widely accepted to be the predominant tools for the structural characterizations of compounds from complex matrixes in many areas. Among these, time-of-flight mass spectrometry and its hybrid or in combination use with tandem mass spectrometry are especially expected to be the most powerful tool for structurally characterizing nontarget compounds, in view of their complementary capacity on providing multistage fragmentations and accurate mass measurements, both of which are important and indispensable information for identifying nontarget compounds using mass spectrometry. Recently, the combination use of LC/IT-MS^n and LC/TOF-MS has been successfully applied to the target identification of major constituents in Radix Salvia miltiorrhizae,[14] diphenhydramine in sediment samples,[16] and metabolites of postharvest fungicides.[17] A very useful four step approach for identifying nontarget components based on the combination use of LC/TOF-MS and LC/IT-MS has been recently developed and successfully applied to several studies.[17−19] In this approach, the chemical formula predicted from the accurate mass was used for database searching for matching structure hits, from which the exact structure could then be determined based on the fragmentation analysis. Theoretically, such an approach is generally applicable to identify any unknown component, provided the corresponding database is comprehensive and accessible. However, it has been observed in our preliminary study that identifying unknown compounds from herbal medicines using this approach was still a difficult task, considering that tens to hundreds of structure hits were retrieved from the database searching by single chemical formula. Too many hits would then make it extremely difficult and time-consuming to locate the exact structure based on the subsequent fragmentation comparisons. Therefore, developing some strategies to narrow the database hits would be very helpful for the nontarget identifications. More recently, Polettini et al. developed such a strategy for characterizing pharmaco/toxicologically relevant compounds (PTRC) in biological samples by creating a subset database from Pubchem Compound.[20] Although this strategy has been proven very useful for PTRC screening, it is unlikely applicable to the herbal components characterizations since it is currently impractical to create such a subset database for herbal components.

This study was thus aimed to develop a generally applicable approach and methodology for the global identification of nontarget components from herbal preparations based on LC/MS-IT-TOF analysis and an original developed strategy. Such a strategy was mainly proposed from an idea that the components contained in herbal preparations can usually be classified into families, and a certain family of components contains common chemical moieties from which the common fragment ions can be produced in tandem mass spectrometry. Following this idea, a strategy involving three steps of procedures including the automated determination of common diagnostic ions and based on family classifications, network establishment and database querying has been developed in this study. With the benefit of the family network, all diagnostic ions can be structurally characterized once a single component in the network has been de novo identified. The characterized structures of the diagnostic ions can then be used as "a priori" information to efficiently select the "exact candidates" that contain such a substructure from the vast primary database hits. The prominent advantage of this strategy can be simply described as transforming the complete "nontarget identification" to "semitarget identification".

Such a novel approach has been successfully applied to the global identification of nontarget components in Mai-Luo-Ning injection (MLN), a well-known herbal prescription which has been widely used in China for more than 30 years for the treatment of cerebral thrombosis, vascular occlusion of angeitides, and deep vein thrombosis of the lower limbs.[21−24] MLN is prepared from the extract mixtures of four herbs, including Flos Lonicerae, Radix Scrophulariae, Radix Achyranthis Bidentatae, and Herba Dendrobii. Up until now, very little is known about their chemical constitutions except several phenolic acids are present.[25] Using the presently developed approach and methodology, we originally detected and identified 87 components from MLN.

EXPERIMENTAL SECTION

Chemicals and Materials. MLN (No. 200611032) was freshly produced in Jinling Pharmaceutical Co., Ltd. (Nanjing, Jiangsu, China). HPLC grade methanol was obtained from Merck (Darmstadt, Germany). The solid phase extraction (SPE) cartridges (MAX, 6 mL, 150 mg; HLB, 6 mL, 200 mg) were purchased from Waters (Milford, MA). All other chemicals were of reagent grade. A Milli-Q Gradient A10 ultrapure water system from Millipore (Milford, MA) was used to obtain the HPLC grade water.

Sample Pretreatment. In order to eliminate the potential interferences from pharmaceutical adjuvant and to make a preseparation of the complicated components, MLN was subjected to a sequential solid phase extraction pretreatment. MLN was diluted 10-fold with ultrapure water and then loaded onto the MAX cartridge which was then eluted with 3 mL of 90% methanol aqueous solution (v/v) containing 5% formic acid. The obtained elution was subsequently loaded onto the HLB cartridge. The MAX cartridge was then sequentially washed with 3 mL of 50% ammonia (v/v) and 3 mL of methanol to elute out the organic acids. The HLB cartridge was washed with 3 mL of 5% methanol aqueous solution (v/v) and eluted with 3 mL of 100% methanol subsequently. Both elution from MAX and HLB cartridges were evaporated to dryness under a gentle stream of nitrogen gas in a water bath. The residues were resuspended in 200 μL of 90% methanol aqueous solution (v/v) containing 5% formic acid and

(16) Ferrer, I.; Heine, C. E.; Thurman, E. M. *Anal. Chem.* **2004**, *76*, 1437–1444.
(17) Thurman, E. M.; Ferrer, I.; Zweigenbaum, J. A.; Garcia-Reyes, J. F.; Woodman, M.; Fernandez-Alba, A. R. *J.Chromatogr., A* **2005**, *1082*, 71–80.
(18) Thurman, E. M.; Ferrer, I.; Fernández-Alba, A. R. *J. Chromatogr., A* **2005**, *1067*, 127–134.
(19) Thurman, E. M.; Ferrer, I.; Heine, C. E. *Anal. Chem.* **2004**, *76*, 1228–1235.
(20) Polettini, A.; Gottardo, R.; Pascali, J. P.; Tagliaro, F. *Anal. Chem.* **2008**, *80*, 3050–3057.
(21) Du, C. J. *Zhongguo Zhongxiyi Jiehe Zazhi* **1996**, *16*, 447–448.
(22) Huang, P. P.; Wang, S. G.; Hua, G. X. *Zhongguo Zhongxiyi Jiehe Zazhi* **1994**, *14*, 86–88.
(23) Wang, X. W.; Hu, H. T.; Dou, W. C. *Zhongguo Zhongxiyi Jiehe Zazhi* **2005**, *25*, 1016–1019.
(24) Wang, X. W.; Hu, H. T.; Xu, J. H.; Yang, Y. X. *Zhongguo Zhongxiyi Jiehe Zazhi* **2005**, *30*, 1549–1551.
(25) Cui, N.; Hao, H.; Wang, G.; Wang, W. *Chromatographia* **2008**, *68*, 33–39.

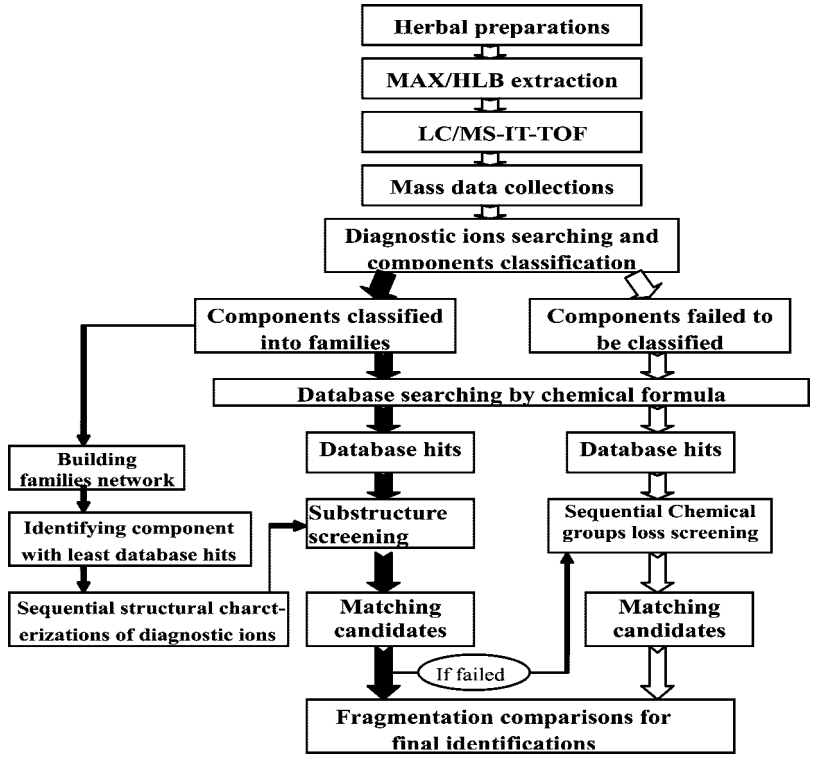

Figure 1. Summary diagram of presently developed approaches and methodologies.

200 μL of methanol, respectively. Following centrifugation (20 000 rpm, 10 min, 2 times), an aliquot of 5 μL was injected into the LC/MS-IT-TOF.

Chromatography and Mass Spectrometry Conditions. All sample analyses were carried out on a LC/MS-IT-TOF system (Shimadzu, Tokyo, Japan). Chromatographic separation was carried out on a column of Synergi C_{18} Hydro-RP 80A, 250 mm × 4.6 mm i.d., 4 μm (Phenomenex). The column oven temperature was set at 35 °C throughout the whole analytical procedure. The mobile phase was delivered at a flow rate of 0.8 mL/min consisting of 0.025% formic acid in water (A) and methanol (B) using a gradient program as follows: 8−12% (B) from 0−15 min, 12−60% (B) from 15−40 min, 60% (B) from 40−50 min, 60−8% (B) from 50−55 min, and 8% (B) from 55−65 min. The mobile phase was split into two identical parts with 0.4 mL/min flowing into the MS detector.

The IT-TOF mass operation parameters were set as follows: negative ion electrospray, nebulizing gas (N_2) flow rate 1.5 L/min, drying gas (N_2) pressure 0.1 MPa, applied probe voltage −3.5 kV, CDL voltage set at constant mode (optimized by autotuning), and CDL temperature 200 °C. Mass spectrometry was conducted in the full scan and automatic multiple stage fragmentation scan modes over an m/z range of 100−1500 for MS and 50−1500 for MS^2 and MS^3 scan, respectively. In the automatic mode, all ions were first accumulated in the octopole and then rapidly pulsed into the IT for MS^n analyses according to the criteria settings. All ions produced were finally introduced into the TOF instrument for accurate mass determination. The ion accumulation time was set at 30 ms. Argon was used as the collision gas, and three different collision energy (30%, 80%, and 150%) induced fragmentations were performed to obtain sufficient fragment ions and to select the most sensitive response for each product ion. Trifluroacetaic acid (TFA) sodium solution was used as the standard

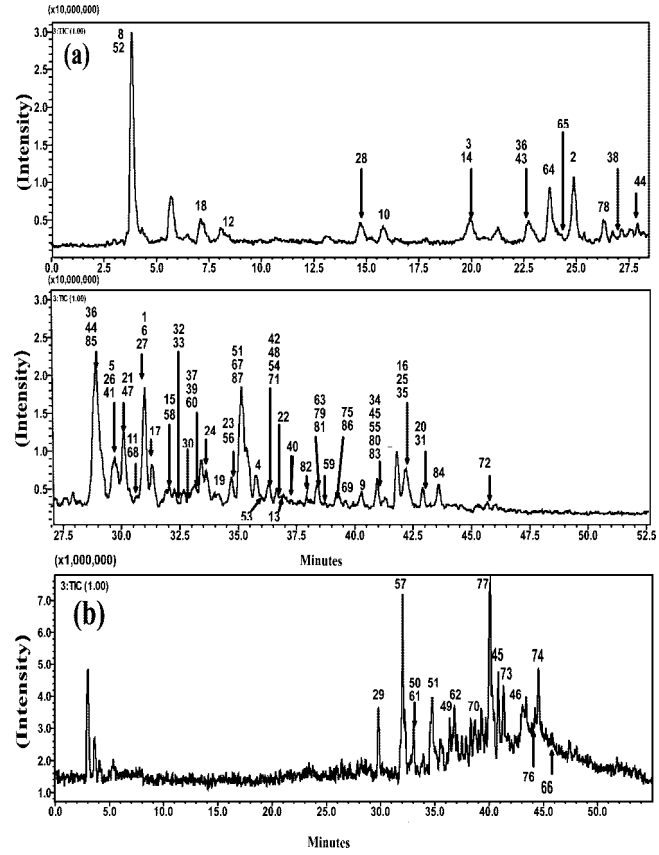

Figure 2. Total ion chromatograms of MLN injection by LC/MS-IT-TOF operated under the negative ionization mode: (a) extracted by MAX cartridge and (b) extracted by HLB cartridge.

sample for calibrating the instrument against the entire mass range (m/z 50−5000). The data recorded were processed by the LC/

Table 1. Retention Time, Accurate Mass, Mass Error, and Chemical Formula of Components in the MNL Injection Detected by LC/MS-IT-TOF

no.	t_R(min)	measured m/z	predicted m/z	error (ppm)	formula	no.	t_R(min)	measured m/z	predicted m/z	error (ppm)	formula
1	30.928	135.0454	135.0452	1.48	$C_8H_8O_2$	45	40.839	377.1816	377.1817	−0.27	$C_{17}H_{30}O_9$
2	24.810	137.0242	137.0244	−2.19	$C_7H_6O_3$	46	43.189	381.2136	381.2130	1.57	$C_{17}H_{34}O_9$
3	19.854	153.0192	153.0193	−0.65	$C_7H_6O_4$	47	30.193	389.1099	389.1089	2.57	$C_{16}H_{22}O_{11}$
4	35.745	163.0404	163.0401	1.84	$C_9H_8O_3$	48	36.390	389.1246	389.1242	1.03	$C_{20}H_{22}O_8$
5	29.748	177.0196	177.0193	1.69	$C_9H_6O_4$	49	36.370	393.1750	393.1766	−4.07	$C_{17}H_{30}O_{10}$
6	31.104	179.0348	179.0350	−1.12	$C_9H_8O_4$	50	32.895	395.1912	395.1923	−3.80	$C_{17}H_{32}O_{10}$
7	41.836	187.0975	187.0976	−0.53	$C_9H_{16}O_4$	51	35.218	403.1250	403.1246	0.99	$C_{17}H_{24}O_{11}$
8	3.815	191.0563	191.0561	1.05	$C_7H_{12}O_6$	52	3.862	405.1030	405.1038	−1.97	$C_{16}H_{22}O_{12}$
9	40.300	193.0501	193.0506	−2.59	$C_{10}H_{10}O_4$	53	35.962	417.1398	417.1402	−0.96	$C_{18}H_{26}O_{11}$
10	15.905	201.0766	201.0768	−0.99	$C_9H_{14}O_5$	54	36.369	429.1395	429.1402	−1.63	$C_{19}H_{26}O_{11}$
11	30.615	210.0773	210.0772	0.48	$C_{10}H_{13}NO_4$	55	40.753	431.0990	431.0984	1.39	$C_{21}H_{20}O_{10}$
12	8.306	215.0559	215.0561	−0.93	$C_9H_{12}O_6$	56	34.883	431.1347	431.1348	−0.23	$C_{22}H_{24}O_9$
13	36.919	217.1083	217.1081	0.92	$C_{10}H_{18}O_5$	57	32.105	431.1938	431.1923	3.48	$C_{20}H_{32}O_{10}$
14	19.710	227.0555	227.0561	−2.64	$C_{10}H_{12}O_6$	58	31.773	433.1347	433.1352	−1.15	$C_{18}H_{26}O_{12}$
15	31.857	241.1203	241.1234	−12.86	$C_{16}H_{18}O_2$	59	38.878	433.1490	433.1504	−3.23	$C_{22}H_{26}O_9$
16	42.197	243.1240	243.1238	0.82	$C_{12}H_{20}O_5$	60	33.075	435.0929	435.0933	−0.92	$C_{20}H_{20}O_{11}$
17	31.430	253.0713	253.0718	−1.98	$C_{12}H_{14}O_6$	61	33.068	435.1517	435.1508	2.07	$C_{18}H_{28}O_{12}$
18	7.263	257.0304	257.0303	0.39	$C_{10}H_{10}O_8$	62	36.610	439.1819	439.1821	−0.46	$C_{18}H_{32}O_{12}$
19	34.110	257.0453	257.0455	−0.78	$C_{14}H_{10}O_5$	63	38.522	447.0924	447.0933	−2.01	$C_{21}H_{20}O_{11}$
20	42.920	263.1284	263.1289	−1.52	$C_{15}H_{20}O_4$	64	23.712	453.1026	453.1038	−2.65	$C_{20}H_{22}O_{12}$
21	30.250	273.0771	273.0768	1.10	$C_{15}H_{14}O_5$	65	24.086	461.1671	461.1675	2.17	$C_{20}H_{30}O_{12}$
22	36.730	279.1229	279.1238	−1.79	$C_{15}H_{20}O_5$	66	45.455	467.2144	467.2134	2.14	$C_{20}H_{36}O_{12}$
23	34.657	281.0663	281.0667	−1.42	$C_{13}H_{14}O_7$	67	35.024	471.1136	471.1144	−1.70	$C_{20}H_{24}O_{13}$
24	33.415	281.1384	281.1394	−3.56	$C_{15}H_{22}O_5$	68	30.756	475.1822	475.1821	0.21	$C_{21}H_{32}O_{12}$
25	42.423	287.0565	287.0561	1.39	$C_{15}H_{12}O_6$	69	39.554	479.2130	479.2134	−0.83	$C_{21}H_{36}O_{12}$
26	29.907	291.0134	291.0146	−4.12	$C_{13}H_8O_8$	70	38.274	481.2650	481.2654	−0.83	$C_{22}H_{42}O_{11}$
27	30.973	293.0666	293.0667	−0.34	$C_{14}H_{14}O_7$	71	36.345	497.1290	497.1301	−2.21	$C_{22}H_{26}O_{13}$
28	14.509	315.0729	315.0722	2.22	$C_{13}H_{16}O_9$	72	45.651	504.1870	504.1875	−0.99	$C_{25}H_{31}NO_{10}$
29	29.733	315.1074	315.1085	−3.49	$C_{14}H_{20}O_8$	73	41.259	507.2080	507.2083	−0.59	$C_{22}H_{36}O_{13}$
30	32.962	325.0932	325.0929	0.92	$C_{15}H_{18}O_8$	74	44.641	507.2951	507.2963	−2.37	$C_{28}H_{44}O_8$
31	42.843	331.1178	331.1187	−2.72	$C_{18}H_{20}O_6$	75	39.244	515.1191	515.1195	−0.19	$C_{25}H_{24}O_{12}$
32	32.235	335.0777	335.0772	1.49	$C_{16}H_{16}O_8$	76	44.224	521.2751	521.2756	−0.77	$C_{28}H_{42}O_9$
33	32.363	337.0926	337.0929	−0.89	$C_{16}H_{18}O_8$	77	40.119	525.3067	525.3069	−0.38	$C_{28}H_{46}O_9$
34	41.008	347.1706	347.1711	−1.44	$C_{16}H_{28}O_8$	78	26.164	531.1140	531.1144	−0.75	$C_{25}H_{24}O_{13}$
35	42.360	349.1865	349.1868	−0.86	$C_{16}H_{30}O_8$	79	38.317	546.1611	546.1617	−1.10	$C_{26}H_{29}NO_{12}$
36	27.445	353.0872	353.0878	−1.42	$C_{16}H_{18}O_9$	80	40.568	577.1549	577.1563	−2.43	$C_{27}H_{30}O_{14}$
37	33.188	353.1028	353.1031	−0.85	$C_{20}H_{18}O_6$	81	38.452	593.1511	593.1512	−0.17	$C_{27}H_{30}O_{15}$
38	27.050	359.0779	359.0772	1.95	$C_{18}H_{16}O_8$	82	37.917	623.1957	623.1981	−3.85	$C_{29}H_{36}O_{15}$
39	33.296	367.1029	367.1035	−1.63	$C_{17}H_{20}O_9$	83	40.938	637.2124	637.2138	−2.20	$C_{30}H_{38}O_{15}$
40	37.282	371.0781	371.0772	2.43	$C_{19}H_{16}O_8$	84	43.374	651.2289	651.2294	−0.77	$C_{31}H_{40}O_{15}$
41	29.599	373.1141	373.1140	0.27	$C_{16}H_{22}O_{10}$	85	28.870	729.1659	729.1672	−1.78	$C_{34}H_{34}O_{18}$
42	36.157	373.1299	373.1293	1.61	$C_{20}H_{22}O_7$	86	39.274	783.2727	783.2717	1.28	$C_{36}H_{48}O_{19}$
43	22.738	375.0943	375.0933	2.60	$C_{15}H_{20}O_{11}$	87	35.174	829.2386	829.2408	−2.65	$C_{36}H_{46}O_{22}$
44	29.292	375.1301	375.1297	1.07	$C_{16}H_{24}O_{10}$						

MS solution software (Shimadzu, Tokyo, Japan), including a formula predictor to predict the chemical formula.

Peak Selections and Data Processing. It has been found in the preliminary study that the peaks with intensity below 100 000 gave few fragments; therefore, only the peaks detected with intensity over 100 000 were selected for identifications. The chemical formulas for all parent and fragment ions of the selected peaks were calculated from the accurate mass using a formula predictor by setting the parameters as follows: C [0−60], H [0−120], O [0−30], N [0−10], double bond equivalent (DBE) [0−20], and H/C ratio [0−3]. Other elements such as P, S, Cl, and Br were not considered since they are rarely present in herbal components. All relevant data including peak number, retention time, accurate mass, the predicted chemical formula, and corresponding mass error were recorded into an Excel file.

Strategy for Nontarget Components Identification. The first step of this strategy is to search for the common fragment ions from all experimentally generated ions, based on a simple program developed by the authors in the Matlab environment (The Mathworks, Natick). The maximum tolerance of mass error was set at 5 mDa when searching for common ions, and only the common ions shared by at least three components were selected out for the purpose of family classification. The components sharing same fragment ions were then classified into a family.

The second step is to build a network of all families through the "bridging components" which are present in two or more families. The family containing the most components was predetermined as the network core, from which other families were connected by bridging components.

The third step is to search by chemical formula against the chemical databases including ChemExper Chemical Directory, Pubchem Compound, and Combined Chemical Dictionary. The database hits by each chemical formula were summed and recorded. Subsequently, the component retrieving the least number of database hits in the core family was subjected to de novo identification by detailed theoretical (all database hits) and experimental fragmentation comparisons, based on the previously reported approach.[18] Once the chemical structure of this component has been determined, the common chemical moiety, i.e, the structure of the diagnostic ion, of this family can be determined

Table 2. Diagnostic Ions Determinations, Family Classifications, Primary Database Hits, and Substructure Selected Candidates[a]

no.	diagnostic ions (m/z)	formula	peaks of each family	no. of primary database hits	no. of substructure selected candidates
1	93.0346	C_6H_6O	1, 2, 4, 8	143, /, 170, 41	41, /, 72, 4
2	109.0295	$C_6H_6O_2$	2,3, 8, 12, 28, 53	40, 45, 41, 32, 11, /	7, 12, 4, 1, 11, /
3	133.0295	$C_8H_6O_2$	5, 9, 25, 32	45, /, 209, 17	1, /, 1,5
4	135.0452	$C_8H_8O_2$	1, 6, 9, 16, 17, 27, 32, 36, 39, 43, 65, 75, 82, 83	143, 116, 227, 35, 30, 12, 11, 22, 14, 8, 4, 16, /, 19	5, 12, 6, 0, 1, 1, 5, 8, 6, 1, 2, 8, /, 7
5	161.0244	$C_9H_6O_3$	9, 17, 25, 32, 83, 84, 86	/, 30, 209, 11, 19, 8, 15	/, 1, 1, 5, 7, 4, 15
6	161.0455	$C_6H_{10}O_5$	29, 34, 45, 46, 47, 49, 50, 51, 53, 57, 66, 68, 69, 73, 82	31, 22, 4, 7, 42, 5, 3, 49, 13, 3, 12, 8, 2, 4, 12	16, 22, 4, 7, 9, 5, 3, 34, 13, 1, 12, 6, 1, /, 11
7	163.0401	$C_9H_8O_3$	4, 21, 23, 30, 33, 38	170, 172, 9, /, 21, 118	4, 0, 3, /, 4, 2
8	165.0557	$C_9H_{10}O_3$	14, 37, 47, 48, 51	16, 130, /, 60, 49	3, 0, /, 1, 5
9	173.0455	$C_7H_{10}O_5$	8, 33, 36, 39, 52, 60, 75	41, 21, /, 14, 12, 25, 16	4, 4, /, 6, 2, 2, 8
10	175.0401	$C_{10}H_8O_3$	63, 64, 81, 84, 86	91, 5, 131, /, 15	1,0,0,/,15
11	179.0350	$C_9H_8O_4$	6, 9, 17, 25, 27, 30, 32, 36, 43, 75, 82, 83, 84	116, 227, 30, 209, 12, 36, 11, 22, 8, 16, /, 19, 8	12,6,1,1,1,1,5,8,1,8,/,7,4
12	179.0561	$C_6H_{12}O_6$	47, 50, 51, 58	/, 3, 49, 24	/, 3, 34, 6
13	191.0561	$C_7H_{12}O_6$	8, 33, 36, 39, 52, 75	41, 21, /, 14, 12, 16	4, 4, /, 6, 2, 8
14	193.0506	$C_{10}H_{10}O_4$	9, 39, 41, 58, 64, 84, 86	/, 14, 30, 24, 5, 8, 15	/, 2, 0, 0, 0, 2, 15
15	329.1242	$C_{15}H_{22}O_8$	68, 83, 84, 86	/, 19, 8, 15	/, 7, 4, 15
16	461.1665	$C_{20}H_{30}O_{12}$	65, 82, 83, 86	4, /, 19, 15	2, /, 7, 15
17	475.1821	$C_{21}H_{32}O_{12}$	68, 83, 84, 86	/, 19, 8, 15	/, 7, 4, 15

[a] "/" indicates the hits or candidate numbers for the bridging components that have been shown elsewhere to avoid duplication.

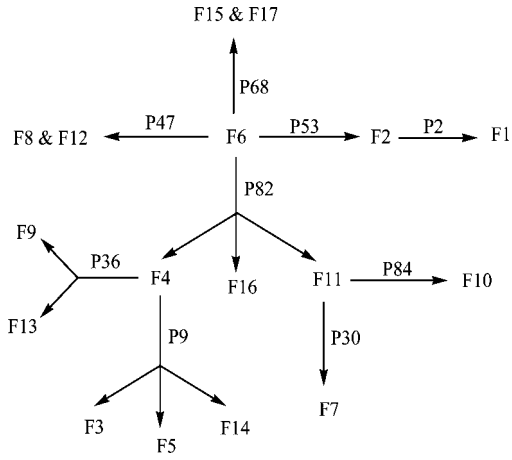

Figure 3. Family network established from bridging components.

from its accurate mass and the fragmentation analysis of this compound. The structurally characterized diagnostic ion can then be used as a very useful "a priori" screening standard for rapidly locating the exact candidates containing such a substructure for all other components in this family. The exact structure of all components in this family could then be determined from these candidates by fragmentation comparisons. After all components in the core family have been identified, the structure of the diagnostic ions for the families directly connected can be readily determined from the identified bridging components. By this rule, all diagnostic ions can be structurally characterized and then used for screening purpose. The general procedures of our strategy and approach are summarized into a diagram as shown in Figure 1.

RESULTS AND DISCUSSION

Peak Detections. Parts a and b of Figure 2 show the TIC chromatogram of the MLN samples pretreated by MAX and HLB solid phase extraction, respectively. The peaks of interest with mass intensity over 100 000 were numbered, and a total of 87 components were detected. The MAX column retained many more components than that by HLB. Furthermore, the mass intensity of most components detected from the MAX extract (× 1 000 000) were much higher than that from HLB (× 100 000). It is well-known that the components contained in herbal medicines usually cover a large range of concentrations (over 10^3 order of magnitude). Therefore, it is necessary to make a preseparation of herbal components before LC/MS analysis to obtain clear and sufficient fragment information for better structural characterizations.

The retention time, accurate mass of deprotonated molecule ion ($[M - H]^-$), mass error, and predicted chemical formula of all detected peaks are summarized in Table 1. For predicting chemical formula, a generally acceptable mass error at 5 ppm was considered for most peaks (86 out of 87), except peak 15 for which the corresponding formula with the least measured error was tentatively selected out.

All deprotonated molecule ions were subjected to up to MS^3 fragmentations. Three collision energies at 30, 80, and 150% were applied to all precursor ions fragmentations in order to produce sufficient product ions with high response. The results are presented in Table S1 (Supporting Information). Although this instrument possesses the capacity of performing up to MS^{10} fragmentations, we conducted only up to MS^3 fragmentation in view that the obtained information is sufficient for structural characterization and the fragmentation over MS^3 is extremely time-consuming for multiple component analysis.

Peak Classifications and Network Establishment. The diagnostic ion based strategy has been previously well proven to be very useful for rapidly identifying targeted compounds.[26,27] However, it is a completely different case in this study for nontarget identification since we know nothing beforehand about the diagnostic ions for nontarget compounds. Herein, we report a novel strategy for determining the diagnostic ions and making classifications for the complete nontarget components. A simple

(26) Zehl, M.; Pittenauer, E.; Jirovetz, L.; Bandhari, P.; Singh, B.; Kaul, V. K.; Rizzi, A.; Allmaier, G. *Anal. Chem.* **2007**, *79*, 8214–8221.

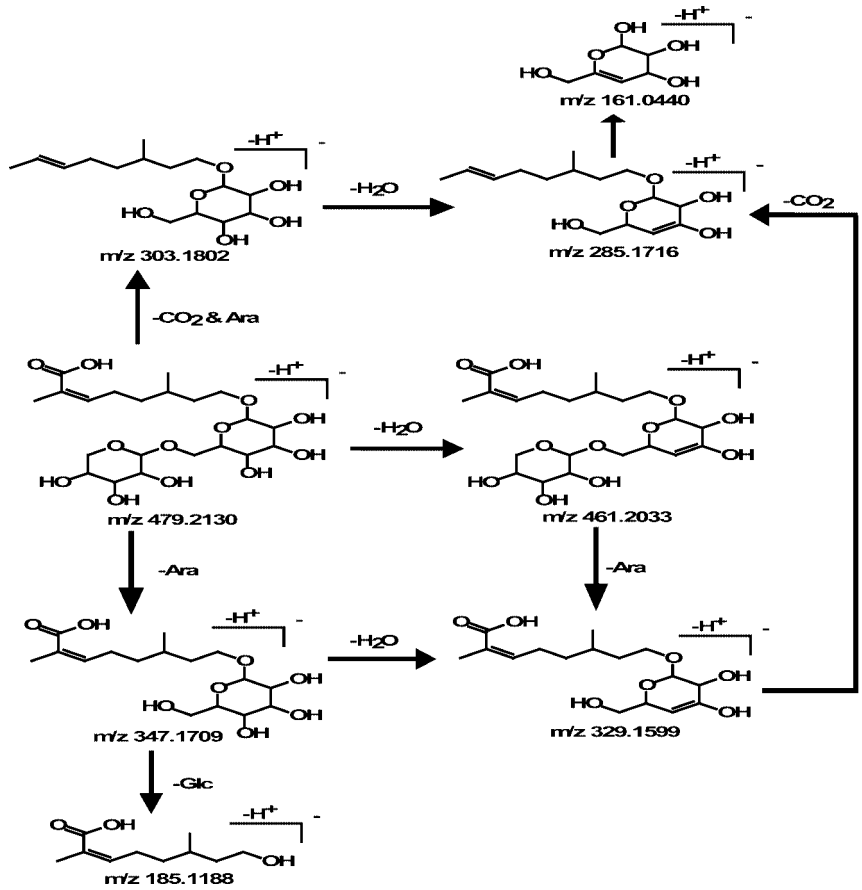

Figure 4. Proposed fragmentations of peak 69.

program was developed in Matlab 7.0 for searching the common ions from all experimentally generated ions of the detected components. As shown in Table 2, 17 diagnostic ions are retrieved by setting the mass tolerance at 5 mDa and selecting out the common ions shared by at least three components. It was observed that 53 out of the total 87 components were successfully classified into families based on the postexperimentally determined diagnostic ions. It was very interesting to find that some peaks are present concurrently in several families. For example, peak 68 is present in families 6, 15, and 17. Therefore, peak 68 was then used as the "bridging peak" for connecting family 6 with family 15 and 17. On the basis of this approach, all families were connected into a coherent network with family 6 (containing the most components) as the network core (Figure 3).

Database Querying and Structure Characterization. Components in family 6 were first subjected to database querying. The summing number of hits was recorded and shown in Table 2. It has been observed that peak 69 has the least number of hits (2 hits) in family 6. Therefore, peak 69 was first subjected to structural characterization by comparing the experimental and theoretical fragmentations. As shown in Figure 4, all experimentally determined fragments match well with that theoretically produced from secolonitoside. The mass errors were within 5 ppm for most ions (except one ion at 9.31 ppm) (Table S2 in the Supporting Information), further supporting the unambiguous identification. Subsequently, the diagnostic ion at m/z 161.0440 for family 6 was readily identified as dehydrate-glucose from the fragmentation analysis of secolonitoside. The structurally characterized diagnostic ion, dehydrate-glucose, was then used as the "a priori" standard to screen the database hits of other components in this family. With the use of such an approach, the primary database hits were substantially narrowed (Table 2). For example, the chemical formula ($C_{16}H_{22}O_{11}$) for peak 47 queried in the databases retrieved 42 hits, while only 9 candidates contain the chemical moiety of dehydrate-glucose. The subsequent identifyication procedures were the same as that for peak 69 by step forward fragmentation comparisons.

After all components in family 6 have been identified, the diagnostic ions for their connecting families could be structurally characterized from the identified bridging components (P68, 47, 53, and 82). As an example, P47 is concurrently present in families 6, 8, and 12. Its identification as theveside readily leads to the structural characterizations of the diagnostic ion of families 8 (m/z, 165.0557, $C_9H_{10}O_3$) and 12 (m/z, 179.0561, $C_6H_{12}O_6$), based on the accurate mass of product ions (Table S3 in the Supporting Information) and the fragmentation analysis (Figure 5). Guided by the family network, all diagnostic ions were structurally characterized as shown in Figure S1 in the Supporting Information. With the use of the structurally characterized diagnostic ions as the "a priori" selecting standard, the average number of exact candidates for all components in the network was only 7 (7 ± 9, range 1−72), nearly 7-fold lower than that of the primary database hits (45 ± 61, range 2−227).

In order to further validate its powerfulness and wide applicability, this strategy has been applied to another known compound herbal preparation, Shengmai injection, which is prepared from the extract mixtures of three herbs including radix ginseng, radix ophiopogonis, and schisandra chinensis. On the

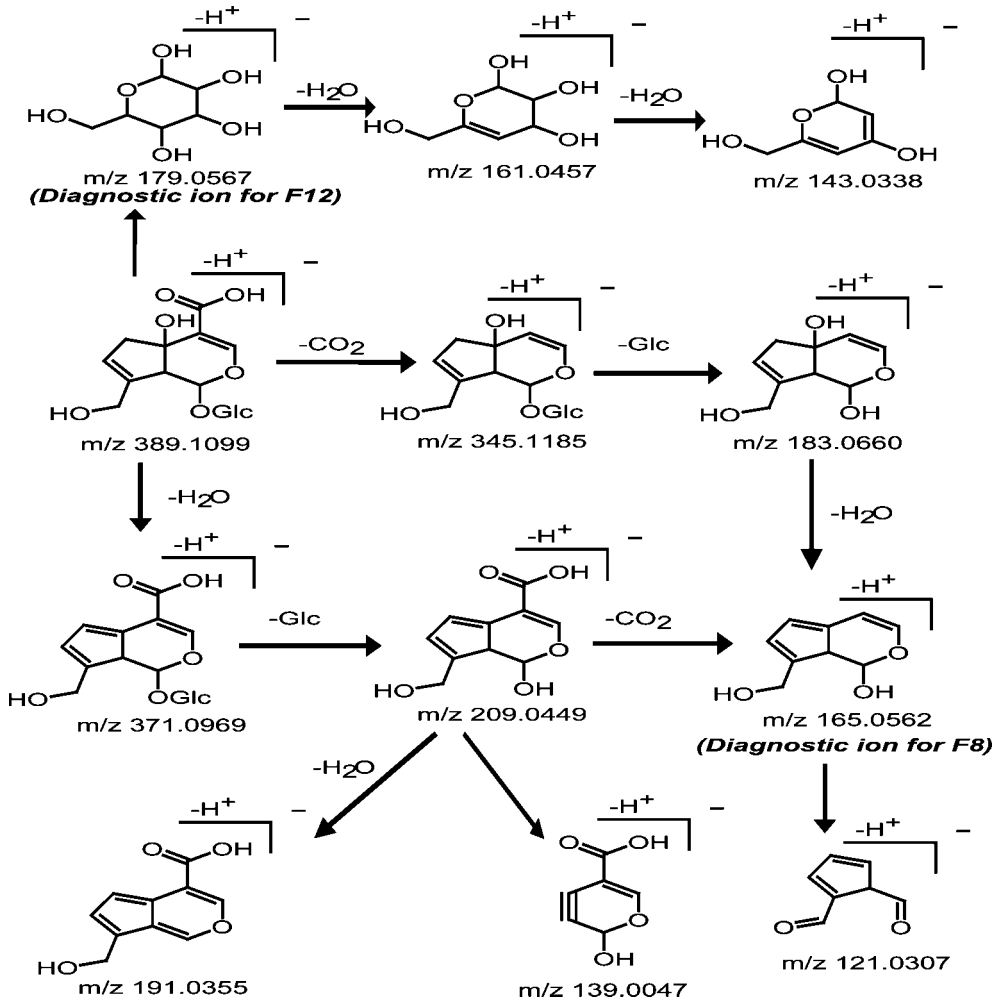

Figure 5. Proposed fragmentations of peak 47 and characterization of the diagnostic ions for families 8 (*m/z*, 165.0562, $C_9H_{10}O_3$) and 12 (*m/z*, 179.0567, $C_6H_{12}O_6$).

basis of the determined 6 diagnostic ions, 41 out of a total of 52 detected peaks were successfully classified into subfamilies and built into two separated networks, one for ginsenosides and another for lignans. All of the 41 peaks have been successfully identified by using this strategy except failed to differentiate some isomers. The identification rate is even higher than that for MLN because of its relatively simpler chemical constitutions, further supporting that this strategy is widely applicable.

Identification of Components That Failed from the Above Strategy. The components that failed to be identified above contain two kinds. One kind is the components (34 components) that failed to be included into the network due to the lack of common diagnostic ions. Another kind is the components (peaks 16, 21, 37, 38, 41, 48, 58, 63, 64, and 81) that retrieved no candidates when screened by their respective diagnostic ion. Such a failure was found to be caused by the proposed structures of the diagnostic ions being incorrect for these 10 components (identical mass corresponding to a different structure). Despite such a mismatch, it would not constitute a serious limitation to applying this strategy into the nontarget identifications. Actually, a success rate of over 80% (43 out of 53) in our study proves that the presently developed strategy is highly useful and powerful for the identifications of nontarget families of compounds.

A complementary approach for relatively rapid screening from the vast list of database hits has been proposed for these 44 components. Taking advantage of the sufficient fragmentations and the accurate mass measurements of fragments, it is possible to directly propose the neutral loss of certain chemical groups from precursor ions to their fragments. In order to facilitate the process of fragmentation comparisons, we developed an approach to sequentially screening the database hits of their capacities on dissociating some specific chemical groups such as H_2O (OH), CO, CO_2 (COOH), benzene, etc. To demonstrate such an approach clearly, we took the peak 19 identification as an example (Mass spectra data is shown in Table S4 in the Supporting Information.). As shown in Figure 6, the first selection of neutral loss H_2O to screen the 99 database hits of peak 19 retrieves 84 candidates, the second by neutral loss of CO retrieves 47 candidates, and the third by further CO loss retrieves 7 candidates, from which peak 19 was tentatively identified as kigelinone based on full fragmentation comparisons.

Summary of Components Identified from MLN. A total of 87 components were detected from MLN injection and all were structurally characterized (Table S1 and Figure S2 in the Supporting Information) except failed to differentiate some isomers (peaks 32, 33, 36, 39, and 75) since they completely shared the

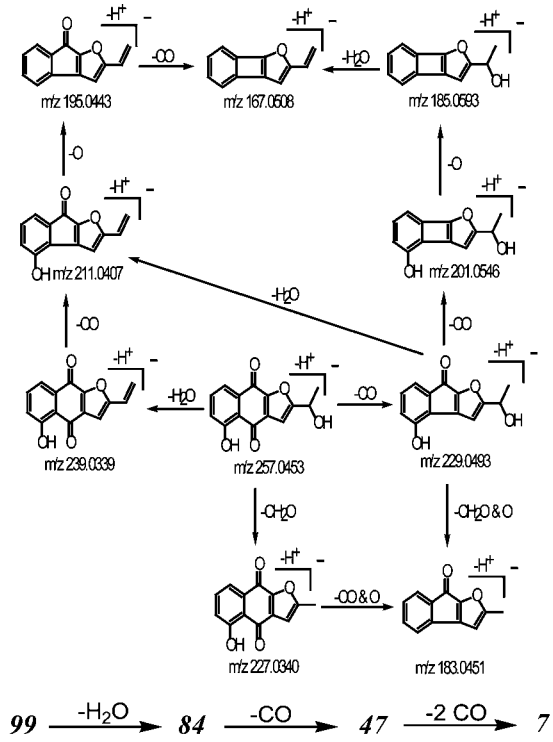

Figure 6. Fragmentations and sequential neutral loss screening procedures for peak 19.

same fragmentations. The identified components can be structurally classified into organic acids (30 components), glycosides (17 components), iridoids (10 components), flavones (10 components), phenylethanoid glycosides (7 components), quinones (4 components), sterones (3 components), alkaloids (3 components), 1 lignan, and 2 others.

CONCLUSION

An approach of database querying by chemical formula combined with fragmentation comparison has been previously proven to be very useful for the identification of nontarget components. However, too many database hits by single chemical formula querying constitute a great difficulty to the subsequent full fragmentation comparison for structure confirmation. This study mainly contributes to a novel strategy for rapid and substantially narrowing the primary database hits based on the diagnostic ion (post experimentally determined) guided family classifications, network establishment, and database screening. Taking MLN as a model herbal preparation, this strategy enables an average 7-fold narrowing of the database hits and thus significantly enhances the efficiency and sharpness for nontarget identifications. A success rate over 80% for identifying the components built into a family network indicates that this strategy is powerful and valuable. The wide applicability of this strategy has been further proven by applying it to another compound herbal preparation. For the components that failed to be identified by this strategy, a complementary approach to screening the database hits by assessing sequential neutral loss of some specific chemical groups has been proposed and proven useful in this study.

In terms of the unequivocal identification of nontarget components, the presently developed strategy and methodology still have some limitations. First, the database querying combined with the fragmentation comparison approach depends largely on the performance and information content of the chemical databases currently available, which means that if the components detected were not included in the targeted chemical databases, it is impossible to identify such components by this approach. Second, the diagnostic ion guided family classification strategy may sometimes fail to incorporate some components when their corresponding diagnostic ions fail to be produced under certain conditions, especially considering that LC/MS spectra are somewhat dependent on the conditions applied. To address such a limitation, the fragmentations should be performed under multiple CID energies to produce sufficient fragments with high response. Third, the inherent limitation of LC/MS based methodology is of that it alone can never suffice for the unequivocal identification of nontarget components. Because of such limitations, we can not exclude the possibility of wrong identifications for some components, especially for those which their real structures are not included in the targeted chemical databases. However, the presently developed strategy and methodology has been well proven to be useful and valuable in this study by successful application to the identification of nontarget components from two compound herbal preparations. Such limitations will not prevent its wide application into the identifications of nontarget components from various complex matrixes.

The identifications of nontarget compounds from complex mixtures are of great significance in many areas such as pharmaceutics,[20] metabonomics,[28] and environmental analysis.[29−31] It is expected that our strategy will find its wide application into not only herbal preparations but to many other complex mixtures such as environmental and biological samples as well, in view that the compounds contained in such mixtures are also structurally related and can be clustered into families.

ACKNOWLEDGMENT

H.H. and N.C. contributed equally to this work. This work was financially supported by the National Natural Science Foundation of China (Grants 30572228, 30630076) and "eleventh-five years" supporting programs from the Ministry of Science and Technology of P.R. China (Grant 2006BAI08B04-05).

SUPPORTING INFORMATION AVAILABLE

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review July 1, 2008. Accepted August 12, 2008.

AC801356S

(27) Sabareesh, V.; Ranganayaki, R. S.; Raghothama, S.; Bopanna, M. P.; Balaram, H.; Srinivasan, M. C.; Balaram, P. *J. Nat. Prod.* **2007**, *70*, 715–729.
(28) Hill, D. W.; Kertesz, T. M.; Fontaine, D.; Friedman, R.; Grant, D. F. *Anal. Chem.* **2008**, *80*, 5574–5582.
(29) Grimalt, S.; Pozo, O. J.; Sancho, J. V.; Hernández, F. *Anal. Chem.* **2007**, *79*, 2833–2843.
(30) Hernandez, F.; Portolés, T.; Pitarch, E.; López, F. J. *Anal. Chem.* **2007**, *79*, 494–504.
(31) Sultan, J.; Gabryelski, W. *Anal. Chem.* **2006**, *78*, 2905–17.