

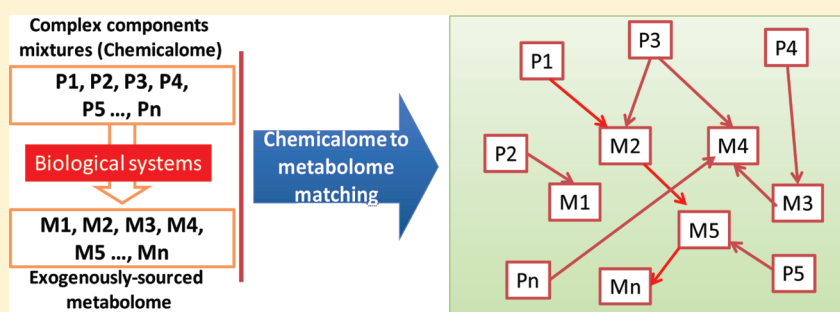
Chemicalome and Metabolome Matching Approach to Elucidating Biological Metabolic Networks of Complex Mixtures

Ping Gong,[†] Nan Cui,[†] Liang Wu,[†] Yan Liang,[†] Kun Hao,[†] Xiangyang Xu,[‡] Weiguo Tang,[‡] Guangji Wang,^{*,†} and Haiping Hao^{*,†}

[†]State Key Laboratory of Natural Medicines, Key Lab of Drug Metabolism and Pharmacokinetics, China Pharmaceutical University, Nanjing 210009, China

[‡]Jinling Pharmaceutical Co., Ltd, Nanjing 210009, China

S Supporting Information



ABSTRACT: Global metabolite identification of complex compound mixtures in biological systems is a very challenging task. Herein, we developed and validated a chemicalome to metabolome matching approach by taking herbal medicine as an example to delineate the metabolic networks of complex systems. This approach consists of five steps of data processing including raw data output, endogenous background subtraction, parent compound and metabolite differentiation, chemicalome to metabolome correlation, and the final validation via manual fragment comparison. Chemicalome to metabolome correlation, the core step of this approach, was performed based on matching the accurate mass differences of pseudomolecular ions between them with the accurate mass changes of known metabolic pathways and validating the matches by validation ions. A step-forward approach that confers a gradual identification of metabolites generated from different steps (1–4) and types (degradation, phase I/II, or mixed) of metabolic reactions was further proposed for chemicalome to metabolome matching. This approach was validated to be very useful and powerful for the metabolite identification of a single compound, a homologous compound mixture, and a complex herbal system. Using this approach, all metabolites (162) detected from urine samples of rats treated with Mai-Luo-Ning injection could be linked to their respective parent compounds, and 143 of them were supported by the final validation via manual fragment analysis. In most cases, more than 80% of the automatic matching results could be supported by the manual fragment validations. A complex metabolic network showing all the possible links between precursors and metabolites was successfully constructed. This study provides a generally applicable approach to global metabolite identification of complex compound mixtures in complex matrixes.

Over the past decade, the appropriateness and robustness of the typical 'one gene, one target, one drug, one disease' drug design paradigm has been questioned because of the gradually decreased productivity of new drug development under this paradigm. Because multiple genes dysregulations are involved in most chronic diseases, it is highly questionable that a drug targeting a single gene will be sufficiently effective in combating such diseases. Network pharmacology-guided multiple-target drug design is becoming an emerging trend in the next stage of drug discovery.¹ It is thus natural to reconsider the ancient concept of botanical/herbal therapeutics that is widely believed to exert therapeutic benefits via a multiple-component and multiple-target mode. Herbal medicine is thereby probably gaining its renaissance through increasing popularity in both academic and industrial fields.² This is indicated by the

increasing number of publications of novel techniques on screening and identifying bioactive compounds from herbal/botanical medicines.^{3–5} However, much work is needed to accomplish herbal medicine-based discovery of multiple-component and multiple-target drugs because of the super-imposed complexities of herbal medicines.

The lack of knowledge in biological dispositions is one of the major obstacles to understanding herbal medicines. Metabolic study, which has long been recognized as an indispensable part in the whole pipeline of drug discovery and development, plays an important role in evaluating drug safety, efficacy, and drug–

Received: January 26, 2012

Accepted: February 21, 2012

Published: February 21, 2012



drug interactions. This also holds true for herbal medicines, in view that the majority of natural compounds are susceptible to biological metabolism. Various natural products such as saponins, quinones, and flavones have been shown to be characterized with extensive metabolism in biological systems.^{6–8} More importantly, it has been found that the metabolites, rather than the parent compounds, are the active species for many natural compounds.^{9,10} For example, the antitumor activity of ginsenoside Rh2 and protopanaxadiol is much stronger than that of the precursor ginsenoside Rg3.^{11,12} In addition, many herbal components are characterized with poor oral bioavailability and extensive metabolism in the gastrointestinal tract where the metabolites produced rather than parent components are more likely to be absorbed into the circulation system.^{13,14} Therefore, the detection and identification of metabolites for herbal compounds is not only a crucial step to uncover the pharmacologically active substances of herbal medicines but may also provide additional resources for discovering new drugs from natural compounds.

However, the detection and identification of metabolites is never an easy task even for single chemical drugs, because of massive endogenous interference and low abundance of metabolites in biological matrices.¹⁵ Although the advance in the performance of high resolution mass spectrometers such as TOF and FT Orbitrap have greatly facilitated the process of metabolite detection and identification,¹⁶ some additional software tools such as the mass defect filter, background subtraction, and ion pattern filtering are necessary for the postacquisition data interpretation.¹⁵ The difficulty in the metabolite identification of complex compound mixtures can thus be conceived. Previous studies concerning metabolite identification of exogenously sourced substances in biological systems were largely limited to isolated single compounds; a few studies concerning complex herbal medicines were largely dependent on the target identification of well-known metabolites of certain types of herbal compounds.^{17–19} We have recently developed an “in vitro to in vivo translation” approach to identifying metabolites of a complex mixture of lignan compounds based on the use of some authentic standards.²⁰ Although this approach is helpful for identifying complicated metabolites of herbal mixtures given that some authentic standards are available, it cannot suffice for the purpose of global metabolite identification of herbal medicines, especially for the nontarget compounds. Global detection and identification of nontarget compounds from complex matrices per se is a very challenging task.^{21,22}

Is it practicable to characterize the global metabolites of nontarget compounds from complex biological matrices? To address this critical question, we propose in this study an original chemicalome–metabolome matching approach (CMMA). Chemicalome is originally defined as the totality of chemicals including drugs, ligands, toxins, and toxic wastes. We defined herein the herbal chemicalome as the totality of intrinsic chemical compounds of herbal medicines. Metabolome originally refers to the complete set of small-molecule metabolites contained in biological systems;²³ we defined herein the herbal metabolome as the global metabolites of the herbal chemicalome produced in biological systems. To facilitate an automatic CMMA, we have developed a computer program that enables rapid and automatic endogenous background subtraction, chemicalome and metabolome differentiation, and step-forward chemicalome to metabolome matching analysis. Using this approach, we have successfully

elucidated the chemicalome to metabolome network of Mai-Luo-Ning (MLN) injection, a compound herbal prescription that we have previously used to develop a ‘diagnostic ion-guided family network approach’ for the global detection and identification of herbal chemicalome.²⁴ Because the present approach is developed for a postacquisition data interpretation, it is independent of the mass spectrometers used and thus is expected to be widely applicable to mass spectrometry-based metabolite identification of various systems of complex components.

■ EXPERIMENTAL SECTION

Chemicals and Reagents. MLN was freshly produced by Jinling Pharmaceutical Co., Ltd. (Nanjing, Jiangsu, China). HPLC grade methanol was obtained from Merck (Darmstadt, Germany). The solid-phase extraction (SPE) cartridges (MAX: 6 cc, 150 mg; HLB: 6 cc, 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 (Billerica, MA) was used to obtain the HPLC grade water.

Animals. Male Sprague–Dawley rats (200–220 g) were obtained from the Academy of Military Medical Sciences (Beijing, China). All the rats were acclimated for at least 1 week before the experiments and allowed water and standard chow ad libitum. All animal studies were approved by the Animal Ethics Committee of China Pharmaceutical University. Rats ($n = 10$) were intravenously injected with MLN at a dose of 10 mL/kg. Rats were kept separately in metabolic cages, and urine samples were collected before dose and after dose at 0–4 h, 4–12 h, and 12–24 h. Urine samples were stored at -20°C before analysis.

Sample Pretreatment. Urine samples diluted with an equal volume of ultrapure water were loaded onto the MAX cartridge which was then washed with 3 mL of 5% ammonia (v/v) and 3 mL of methanol; the obtained washout was subsequently loaded onto the HLB cartridge. The postwash MAX cartridge was eluted with 3 mL of 90% methanol aqueous solution containing 5% formic acid, while the HLB cartridge was washed with 3 mL of 5% methanol solution and then eluted with 3 mL of 100% methanol. Both eluents from MAX and HLB cartridges were evaporated to dryness under a gentle stream of nitrogen gas in a water bath. The obtained residues were resuspended in 200 μL of 90% methanol aqueous solution (v/v) containing 5% formic acid (for MAX elution) and 200 μL of 100% methanol (for HLB elution). The resuspensions were centrifuged two times at 10 000 g for 10 min, and an aliquot of 5 μL was injected into the LC/MS-IT-TOF for analysis.

Chromatography and Mass Spectrometry Conditions. All sample analyses were performed on a LC/MS-IT-TOF system (Shimadzu, Tokyo, Japan). The chromatographic separation and mass spectrometer detection conditions applied in this study are shown in the Supporting Information and were mostly the same as that used in our previous study.²⁴

Chemicalome to Metabolome Matching Approach. Detailed procedures of raw data processing, endogenous interference subtraction, and chemicalome to metabolome differentiation are shown in the Supporting Information. The whole process of chemicalome to metabolome matching consists of the following five steps: (a) outputting raw chromatographic and mass data of all samples based on setting criteria; (b) subtracting endogenous interferences and constructing the exogenously sourced compound data matrixes (ERC file); (c) differentiating chemicalome and metabolome

by comparing the ERC file with the file that records herbal parent compounds (herbal file); (d) matching the mass differences of pseudomolecular ions between the metabolites and parent compounds to the accurate mass changes of typical metabolic reactions, and simultaneously validating the matches based on the search of validation ions; (e) validating the automatic predictions via manual fragment comparison. To facilitate a chemicalome to metabolome matching analysis, a step-forward approach that considers the steps (1–4) and types (degradation, phase I, and phase II) of metabolic reactions was proposed. The scheme for the whole process of chemicalome to metabolome matching analysis is depicted in Figure 1. After

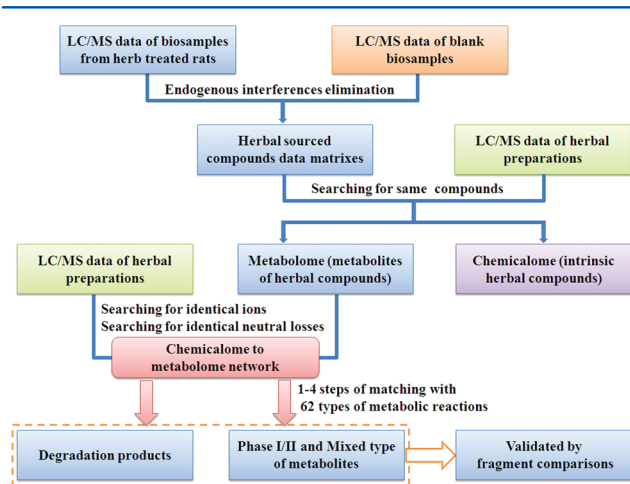


Figure 1. Workflow for “chemicalome to metabolome matching network” approach to the metabolite identification of complex compound mixtures. The whole process is performed by using an in-house-generated computer software “chem-metab explorer”. Of note, the searching of validated ions and the matching with metabolic reactions can be performed simultaneously in the program. The automatic predictions were finally validated by manual fragment comparison. Chemical structures of metabolites were tentatively proposed after their parent compounds in chemicalome and the respective metabolic reactions were clarified from CMMA analysis.

the output of raw data from the LC/MS-IT-TOF workstation, all the processes except the final step of manual fragment comparison to confirm the prediction results were performed automatically in an in-house designed software “Chem-metab explorer” (the program is free accessible in the Supporting Information).

This approach was conceived based on two generic principles; first, the metabolic pathways for exogenous compounds in animals and humans are largely known and the types of metabolic reactions are limited; second, the fragment patterns between metabolites and precursors are similar or even the same. Thus, we incorporated a total of 62 types of known metabolic pathways (Table S1, Supporting Information) into the “chem-metab explorer” software. The mass differences between the pseudomolecular ions of the metabolites and the chemicalome components were calculated and matched with the accurate mass change of typical metabolic reactions based on a step-forward matching process. Two types of “validation ions” between metabolites and precursors were concomitantly searched to narrow and/or validate the matches. The first type is of the identical fragment ions between metabolites and precursors; the second type indicates a pair of ions between metabolites and precursors with

identical accurate mass difference to their respective pseudomolecular ions (identical neutral losses). These two types of “validation ions” (VI) were recorded automatically in “chem-metab explorer”. Considering that the absolute numbers of ions produced among different compounds are varied, we proposed another index, match ratio of ions (MR), to normalize the differences. MR value is expressed as the ratio of two times the validation ions to the total numbers of fragment ions of a metabolite and its possible precursor. To ensure the suitability and correctness of the automatic matching results, appropriate VI numbers and MR values were gradually optimized based on checking the number of automatic predictions for each metabolite. This step-forward process decreases the number of false matching and thus increases the sharpness and correctness of chemicalome to metabolome matching. The automatic predictions were then validated by manual fragment comparison. The performance of automatic matching in predicting various types and/or steps of metabolites was evaluated by an index “validation ratio” that indicates the averaged ratio of automatic predictions for certain types of metabolites validated to be correct by manual fragment analysis. The structure of chemicalome compounds in MLN had been characterized in our previous study;²⁴ once the matching metabolic relationship between metabolites and certain chemicalome compounds were validated using the presently developed CMMA, the structure of the metabolites could then be tentatively proposed from that of the corresponding parent compound and the proposed metabolic pathway.

RESULTS AND DISCUSSION

To validate the usefulness and powerfulness of CMMA, three kinds of herbal samples/data with different extents of complexity were tested. First, we analyzed the data of a single herbal compound, echinacoside, for which the metabolites had been identified manually by us with a step-forward approach.²⁵ Second, metabolite identification of a homologous mixture of compounds, Schisandra lignan extract (SLE),²⁰ was performed using the presently developed approach. Finally, the approach was applied to the analysis of the complicated chemicalome to metabolome network of a compound herbal prescription, MLN.²⁴

Validated Analysis of Previously Published Data.

Echinacoside is a natural compound that undergoes extensive metabolism in the gastrointestinal tract and liver. Because echinacoside involves multiple types and multiple steps of metabolic reactions, the manual identification of its metabolites is not an easy task; the manual process was time-consuming and largely dependent on experienced knowledge of metabolism. To validate the performance and reliability of the presently developed approach, we input our previous data sets into “chem-metab explorer” for the automatic characterization of detected metabolites. The degradation products searching found four matches; the one-step phase I and phase II searching retrieved five peaks, and the remaining four peaks were successfully characterized by a multiple-step metabolism searching. Of note, the direct precursor for the multiple-step metabolite search could be automatically located (Table S2, Supporting Information). In addition, the recorded MR values all exceeded 0.2, supporting that it is practicable to search identical fragment modes between metabolites and precursors. The whole process consumed only around 10 min, and the retrieved results were the same as those of our previous report

by manual analysis. Furthermore, the software produced a sole candidate pathway for each metabolite (i.e., correct in all cases). These results demonstrate that the presently developed approach is highly useful for the identification of drug metabolites.

To further validate the performance of our approach, the data of the metabolite identification of a multiple compound mixture, SLE, were reanalyzed using the presently developed approach. SLE contains dozens of lignan compounds, and thus it is a difficult task to identify their metabolites in biosamples. Our previously developed approach was successful but largely dependent on labor-intensive experimental protocols and manual data processing. By using the presently developed CMMA, a total of 44 metabolites detected from rat urine samples were rapidly and automatically characterized, and the results obtained were in accordance with our previous publication.²⁰ The one-step and two-step metabolism searching retrieved 25 and 19 metabolites, respectively. The validation ratio for the automatic identification of one- and two-step metabolites is $96.2\% \pm 19.6\%$ and $94.2\% \pm 10.1\%$, respectively. Some false candidates were retrieved from the automatic matching; however, it was easy to exclude such false candidates by the manual fragment comparison. This validation result strongly suggests that CMMA is highly efficient and powerful for the metabolite identification of complex compound mixtures.

Chemicalome to Metabolome Matching Analysis of MLN. After the above validation analysis, we next proceeded to test the performance of this approach in the analysis of a much more complex herbal system, MLN. We have previously clarified the chemicalome of MLN injection, from which a total of 87 chemical components that can be classified to 8 chemical families were characterized based on a “diagnostic ion-guided network bridging” approach.²⁴ These herbal components may produce hundreds of metabolites in the biological systems, and thus it is somewhat impractical to identify such a complicated metabolic network based on the previously published approaches for the metabolite identification of single compounds. By using the presently developed CMMA, we have detected and tentatively characterized a total of 180 compounds from the rat urine samples collected between 0 and 24 h after MLN dosing. Among these components, 18 were proposed as parent compounds, 12 as degradation products, 51 as one-step metabolites, 70 as two-step metabolites, and 41 as multiple-step and mixed-type metabolites (Figure S1, Supporting Information).

After the process of background subtraction and chemicalome to metabolome differentiation (detailed results are shown in Supporting Information), the chemicalome to metabolome matching analysis was performed between the retrieved 162 metabolites and 87 parent compounds. To clarify such a complicated chemicalome to metabolome network, we proposed a step-forward approach to the gradual identification of degradation products, one-step phase I and phase II metabolites, two-step metabolites, and multiple-step metabolites. Detailed results for the identification of metabolome and the corresponding metabolic reactions and pathways are shown in Table S4, Supporting Information.

Degradation products were identified based on the “fragmentation–degradation” relationship which indicated that the susceptibility of breaking chemical bonds is similar in mass spectrometers and biological systems.²⁶ On the basis of this rule, we searched the pseudomolecular ions of all the

metabolites against the fragment data sets in the herbal file to find the matches with identical mass values. The located matches were then subjected to an identical fragment ion search and the final manual fragment comparison, from which the degradation products could be confirmed. By this process, 12 degradation products were identified. It is important to note that a certain degradation product can be produced from various parent compounds. Taking peak 6 as an example, we illustrated that this degradation product identified as caffeic acid could be produced from a total of 10 different parent compounds via either degradation or phase I metabolism (Figure S2, Supporting Information). This result clearly indicates the complexity for the metabolite identification of herbal compounds mixtures.

After the identification of degradation products, we next sought to identify the metabolites that can be explained by a single step of metabolic reaction from the compounds recorded in the herbal file. The mass differences of the pseudomolecular ion of each metabolite to those of all the chemicalome compounds were automatically matched with the mass change of one-step metabolic addition or subtraction of certain chemical groups recorded in Table S1. Validation ions that also match the metabolic reactions were simultaneously screened, and the VI numbers and MR values were automatically recorded. We took the identification of M123 as an example to illustrate the whole process of chemicalome to metabolome matching analysis (Figure 2). The accurate mass differences between M123 and 87 parent components were automatically calculated, and 7 of them were found matched with the accurate mass change of known metabolic reactions. Four of the eight candidates characterized with low MR values were excluded. The remaining four candidates were subjected to manual fragment analysis, which indicated that M123 was produced from P17 via one-step of glucuronidation. The MR value over 0.2, which was optimized from a stepwise preliminary run in the program to retrieve an appropriate number of automatic predictions for each metabolite, was set as a condition for the one-step chemicalome to metabolome matching analysis. With one-step metabolism matching, 32 phase I metabolites and 21 phase II metabolites were identified. The final step of manual fragment analysis indicates that the validation ratio of automatic predicting results is very high; the averaged validation ratio for the identification of one-step phase I and phase II metabolites was 92.9% and 84.8%, respectively (Table 1). The identified phase I metabolites contain 4 hydroxylation, 4 hydroxyl reduction, 3 dehydrogenation, 8 oxidation, 7 reduction, and 5 deetherization products; phase II metabolites contain 16 glucuronidation, 3 glycine conjugation, and 4 methylation products (Table S4).

Two-step metabolism indicates that the metabolites are produced from the parent compounds via two-step metabolic reactions. Considering that for some two-step metabolites their direct precursors may have been identified as degradation products or one-step metabolites in the above processes, we performed the matching analysis of the remaining unidentified metabolites with both the herbal chemicalome (two-step matching) and the identified one-step metabolite data sets (one-step matching). As a result, a total of 70 metabolites were found in this process and 19 of them had their direct precursors in the one-step metabolites (Table 1). These metabolites consist of 1 degradation adding phase I, 25 two-step of phase I, 29 typical phase I adding phase II, and 15 two-step of phase II metabolic reactions. The validation ratio of automatic matching

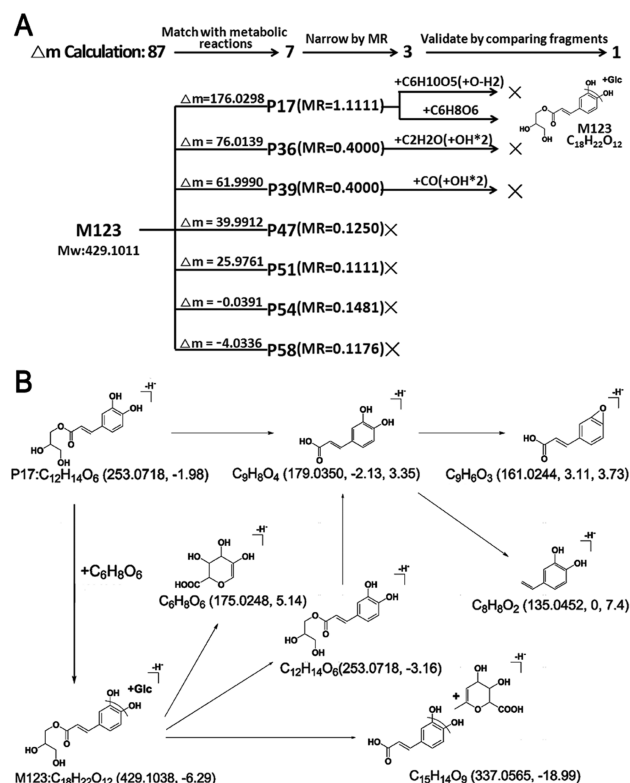


Figure 2. The identification of M123 as a representative example to illustrate the typical process of chemicalome to metabolome matching analysis. A, the whole process of chemicalome to metabolome matching analysis to characterize M123; $\Delta m = m/z$ value of metabolite – m/z value of parent compound. B, fragment comparison of M123 and its parent compound P17 (1-O-cafeoylglycerol) to validate the automatic matching result; data shown in brackets are the predicted mass and the errors (ppm) of experimentally determined mass to the predicted value. Two errors are shown for fragments detected from both parent compound (the former) and metabolite (the latter).

was satisfactory, characterized with the lowest at 55.6% for the two-step matching of sequential phase II metabolites (Table 1). For most of the metabolite matching, the software presented no more than two automatic predictions which could then be readily identified by manual fragment comparison. We took the characterization of the metabolic pathways of parent compound 75 (di-O-cafeoylquinic acid) as an example to show the process of identifying sequential metabolites using the presently developed approach (Figure 3). P75 was first metabolized via methylation to produce both the single methylated (M153) and the double methylated metabolite (M157). With two-step metabolism matching, we found that M174 is the direct glucuronidated product of M157 and thus the two-step metabolite of P75. As clearly shown in both Figures 2 and 3, the presently developed CMMA can rapidly delineate the metabolic relationships; however, it is difficult to determine the exact metabolic site when multiple possibilities are available.

Figures 2 and 3 show that some predicted metabolites can also be found in the herbal preparations and are indeed the parent compounds. This result was not a surprise considering that the herbal components are produced via sequential metabolism in the plants and that the plants and animals share common machineries in metabolizing natural compounds. In these cases, it is impractical to ascertain the exact contributions of each parent compound to producing the final metabolites without the quantitative determination of the parent compounds and metabolites involved. However, we consider that this phenomenon does not hamper the application of our approach to delineate the metabolic network of complicated herbal components, and, to the contrary, it reflects the power of our approach to constructing the complicated network of herbal chemicalome to metabolome.

After the above steps of metabolite searching, there were still 41 metabolites remaining unidentified, indicating that there might have been some metabolites produced via multiple-step metabolism. Our previous studies also supported that multiple-step metabolism of natural products was not a rare phenomenon.^{6,20,25,27} To this end, multiple-step matching

Table 1. General Information about the Metabolite Identification of MLN Injection Using the Chemicalome to Metabolome Matching Approach^a

steps	metabolic types	numbers of metabolites	averaged automatic predictions	validation ratio (%)	VIN	MR
0	prototype	18	1.00 ± 0.00	100 ± 0.00	3.00 ± 2.38	0.55 ± 0.26
	degradation	12	1.33 ± 0.82	100 ± 0.00	1.92 ± 1.00	0.37 ± 0.15
1	phase II	21	1.09 ± 0.29	84.8 ± 35.1	2.50 ± 1.97	0.46 ± 0.26
	phase I	32	1.29 ± 0.60	92.9 ± 26.2	2.25 ± 1.21	0.44 ± 0.22
1 + 1	II + II	8	1.22 ± 0.44	94.7 ± 10.9	1.75 ± 0.46	0.46 ± 0.17
	I + I	3	1.40 ± 0.89	100 ± 0.00	2.00 ± 0.00	0.36 ± 0.19
	I + II	7	1.36 ± 0.92	81.8 ± 40.5	2.29 ± 1.11	0.43 ± 0.14
	I + degradation	1	1.00 ± 0.00	100 ± 0.00	1.00 ± 0.00	0.20 ± 0.00
2	II + II	7	1.11 ± 0.33	55.6 ± 52.7	1.43 ± 0.79	0.29 ± 0.10
	I + I	22	2.00 ± 1.63	68.3 ± 37.2	1.68 ± 0.95	0.33 ± 0.19
	I + II	22	2.47 ± 1.95	62.5 ± 25.0	1.58 ± 0.67	0.30 ± 0.12
2 + 1	I + II	14	1.44 ± 0.73	88.9 ± 33.3	2.21 ± 2.33	0.38 ± 0.16
	II + degradation	2	1.00 ± 0.00	100 ± 0.00	1.50 ± 0.58	0.26 ± 0.06
3	I + II	17	3.47 ± 2.65	58.2 ± 38.9	1.47 ± 0.62	0.29 ± 0.12
3 + 1	I + II	2	1.40 ± 0.89	100 ± 0.00	1.57 ± 0.53	0.29 ± 0.16

^aData shown are means ± SD. VIN, numbers of validated ions; MR, match ratio, calculated from the equation: $MR = VIN \times 2 / (PFN + MFN)$, where PFN and MFN indicate the numbers of fragment ions of a certain parent compound and metabolite, respectively. Validation ratio is calculated by the total numbers of automatic predictions divided by the numbers that are validated to be correct by manual fragment analysis. Metabolic steps of 1 + 1, 2 + 1, and 3 + 1 represent two separated running of match and indicate that the finally retrieved metabolites could find their direct precursors in the former identified metabolite data sets.

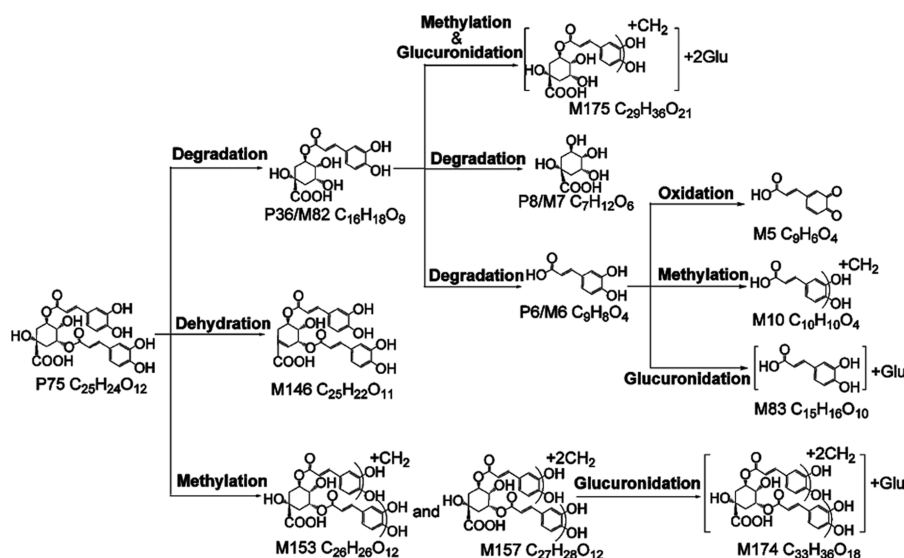


Figure 3. Full metabolic pathways of P75 (di-*O*-caffeoylquinic acid) contained in MLN injection. This typical example is to show that the chemicalome to metabolome matching approach is able to characterize the complete and sequential metabolic pathways of herbal compounds in complex mixtures.

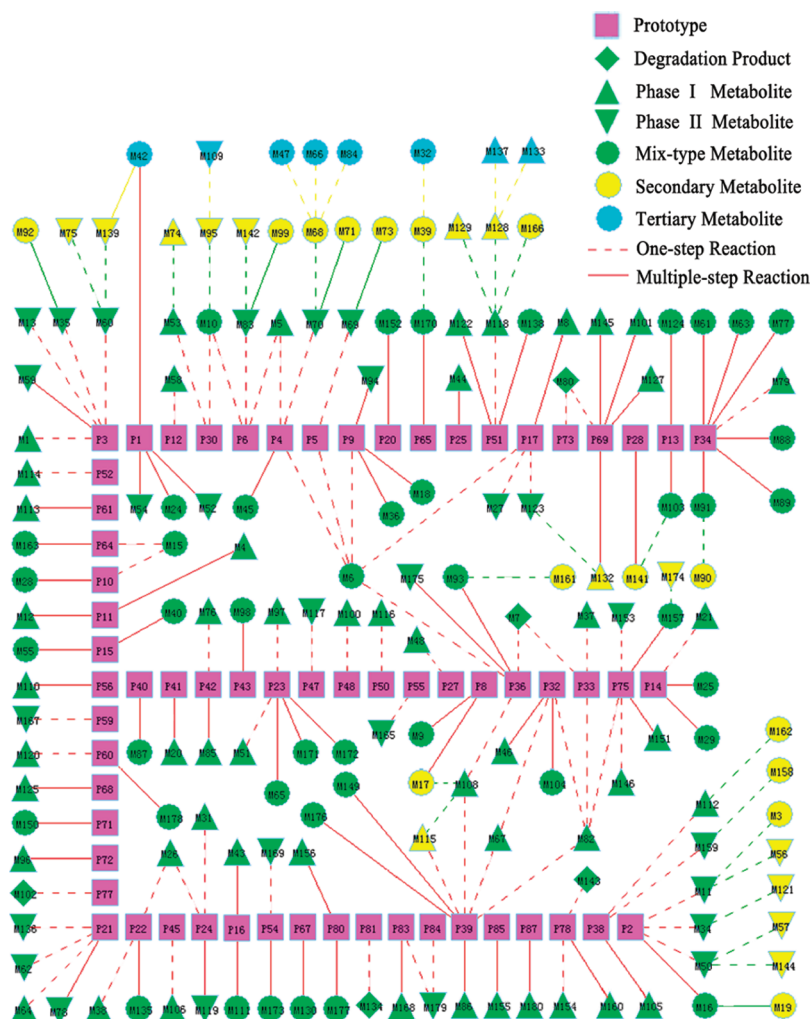


Figure 4. Chemicalome to metabolome matching network of MLN injection.

was performed for the identification of the remaining 41 metabolites. As a result, all metabolites had predicted

candidates from the automatic multiple-step matching, and all of them were characterized with at least one pair of validation

ions to support the prediction results. However, only 22 of them could be supported by the manual analysis of fragments. The validation ratio for the multiple-step metabolites was much lower than that for the one- and two-step metabolites. We found that it was extremely difficult to compare and correlate the fragment ions between the parent compound and some multiple-step metabolites for which the direct precursors were not detected in the samples. In some cases, the fragment information was not sufficient, possibly because of the low abundance of such metabolites, to support the validation of automatic matching candidates. It is also possible that for such unidentified metabolites their real metabolic reactions have not been included into the program to match. These factors all add to the uncertainties in the identification of multiple-step metabolites for herbal compound mixtures.

As discussed in our previous publication,²⁴ mass spectrometry alone without the validation using authentic standards can never suffice for the unequivocal identification of nontarget components and thus we could not exclude the possibility of wrong identifications of some chemicalome components in MLN injection. Such a limitation will certainly lead to further uncertainty in characterizing the exact chemical structures of the metabolites of such compounds. In addition, because of the intrinsic limitation of mass spectrometry-based metabolite identification, it is quite difficult to determine the exact metabolic sites when multiple possibilities are available. Despite these limitations, the presently developed CMMA represents a powerful approach to delineate the metabolic pathways of complex nontarget compound mixtures and to uncover the complicated network relationship between chemicalome and metabolome.

CONCLUSIONS

Biological systems are constantly exposed to various exogenously sourced components such as drugs, herbs, food supplements, and environmental contaminants; metabolite identification is thus an important task in various fields. It is noteworthy that biological systems are in most cases exposed to exogenously sourced compound mixtures rather than a single compound; the metabolite identification of such compound mixtures represents a very challenging task. The present study contributes to the development and validation of a “chemicalome to metabolome matching” approach that is expected to be universally applicable to the metabolic studies of various complex mixtures. Our approach has been validated to be highly efficient and reliable in the metabolite identification of a single compound, a homologous compound mixture, and a compound herbal preparation. The metabolic network of MLN was successfully established with this approach (Figure 4). In view that various types of exogenous compounds share common metabolic pathways in the biological systems, the presently developed CMMA will not be limited in the field of herbal analysis and can be expected to be used for the metabolite identification of many other exogenous components such as chemically synthesized drugs, food supplements, and environmental contaminants.

ASSOCIATED CONTENT

Supporting Information

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

AUTHOR INFORMATION

Corresponding Author

*E-mail: hph_770505@yahoo.com.cn, fax: +86-25-83271060 (H.H); e-mail: guangjiwang@hotmail.com, fax: +86-25-83302827 (G.W.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was financially supported by National Key new drug innovative program (Grant 2011ZX09201-201-03), National Basic Research Program of China (“973 Program”) (Grant 2011CB505300-03), a Foundation for the Author of National Excellent Doctoral Dissertation of China (Grant 200979), Natural Science Foundation of Jiangsu province (Grant BK2011065), and the Program for New Century Excellent Talents in University (Grant NCET-09-0770). The first two authors contributed equally to this work.

REFERENCES

- (1) Hopkins, A. L. *Nat Chem Biol* **2008**, *4*, 682–690.
- (2) Schmidt, B. M.; Ribnicky, D. M.; Lipsky, P. E.; Raskin, I. *Nat. Chem. Biol.* **2007**, *3*, 360–366.
- (3) Fu, H.; Fu, W.; Sun, M.; Shou, Q.; Zhai, Y.; Cheng, H.; Teng, L.; Mou, X.; Li, Y.; Wan, S.; Zhang, S.; Xu, Q.; Zhang, X.; Wang, J.; Zhu, J.; Wang, X.; Xu, X.; Lv, G.; Jin, L.; Guo, W.; Ke, Y. *Anal. Chem.* **2011**, *83*, 6518–6526.
- (4) Zhang, X.; Jia, B.; Huang, K.; Hu, B.; Chen, R.; Chen, H. *Anal. Chem.* **2011**, *82*, 8060–8070.
- (5) Chau, F. T.; Chan, H. Y.; Cheung, C. Y.; Xu, C. J.; Liang, Y.; Kvalheim, O. M. *Anal. Chem.* **2009**, *81*, 7217–7225.
- (6) Hao, H.; Wang, G.; Cui, N.; Li, J.; Xie, L.; Ding, Z. *Curr. Drug Metab.* **2007**, *8*, 137–149.
- (7) Hao, H.; Lai, L.; Zheng, C.; Wang, Q.; Yu, G.; Zhou, X.; Wu, L.; Gong, P.; Wang, G. *Drug Metab. Dispos.* **2010**, *38*, 1731–1739.
- (8) Wang, Q.; Hao, H.; Zhu, X.; Yu, G.; Lai, L.; Liu, Y.; Wang, Y.; Jiang, S.; Wang, G. *Drug Metab. Dispos.* **2010**, *38*, 1132–1140.
- (9) Chen, X. W.; Serag, E. S.; Sneed, K. B.; Zhou, S. F. *Chem. Biol. Interact.* **2011**, *192*, 161–176.
- (10) Hao, H.; Zheng, X.; Li, P.; Wang, G. *Curr. Drug Metab.* **2011**, *12*, 850–870.
- (11) Wang, W.; Wang, H.; Rayburn, E. R.; Zhao, Y.; Hill, D. L.; Zhang, R. *Br. J. Cancer* **2008**, *98*, 792–802.
- (12) Kim, H. S.; Lee, E. H.; Ko, S. R.; Choi, K. J.; Park, J. H.; Im, D. S. *Arch. Pharm. Res.* **2004**, *27*, 429–435.
- (13) Liu, Y. T.; Hao, H. P.; Xie, H. G.; Lai, L.; Wang, Q.; Liu, C. X.; Wang, G. J. *Drug Metab. Dispos.* **2010**, *38*, 1779–1784.
- (14) Kang, A.; Hao, H.; Zheng, X.; Liang, Y.; Xie, Y.; Xie, T.; Dai, C.; Zhao, Q.; Wu, X.; Xie, L.; Wang, G. *J. Neuroinflammation* **2011**, *8*, 100.
- (15) Ma, S.; Chowdhury, S. K. *Anal. Chem.* **2011**, *83*, 5028–5036.
- (16) Martinez Vidal, J. L.; Plaza-Bolanos, P.; Romero-Gonzalez, R.; Garrido Frenich, A. *J Chromatogr A* **2009**, *1216*, 6767–6788.
- (17) Xiang, C.; Qiao, X.; Wang, Q.; Li, R.; Miao, W.; Guo, D.; Ye, M. *Drug Metab. Dispos.* **2011**, *39*, 1597–1608.
- (18) Xin, G. Z.; Qi, L. W.; Shi, Z. Q.; Li, P.; Hao, H. P.; Wang, G. J.; Shang, J. *Curr. Drug Metab.* **2011**, *12*, 809–817.
- (19) Cao, G.; Cai, H.; Zhang, Y.; Cong, X.; Zhang, C.; Cai, B. J. *Pharm. Biomed. Anal.* **2011**, *56*, 118–125.
- (20) Liang, Y.; Hao, H.; Xie, L.; Kang, A.; Xie, T.; Zheng, X.; Dai, C.; Hao, K.; Sheng, L.; Wang, G. *Drug Metab. Dispos.* **2010**, *38*, 1747–1759.
- (21) Schymanski, E. L.; Bataineh, M.; Goss, K. U.; Brack, W. *TrAC Trends Anal. Chem.* **2009**, *28*, 550–561.
- (22) Bataineh, M.; Lubcke-von Varel, U.; Hayen, H.; Brack, W. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1016–1027.
- (23) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.

- (24) Hao, H.; Cui, N.; Wang, G.; Xiang, B.; Liang, Y.; Xu, X.; Zhang, H.; Yang, J.; Zheng, C.; Wu, L.; Gong, P.; Wang, W. *Anal. Chem.* **2008**, *80*, 8187–8194.
- (25) Wang, Y.; Hao, H.; Wang, G.; Tu, P.; Jiang, Y.; Liang, Y.; Dai, L.; Yang, H.; Lai, L.; Zheng, C.; Wang, Q.; Cui, N.; Liu, Y. *Talanta* **2009**, *80*, 572–580.
- (26) Garcia-Reyes, J. F.; Molina-Diaz, A.; Fernandez-Alba, A. R. *Anal. Chem.* **2007**, *79*, 307–321.
- (27) Liu, Y.; Hao, H.; Xie, H.; Lv, H.; Liu, C.; Wang, G. *J. Pharm. Sci.* **2009**, *98*, 4391–4401.