**Annex 3**

**Medcheck: a novel software for automatic de-formulation of Traditional Chinese medicine (TCM) prescriptions by Liquid Chromatography-Mass Spectrometry**

**Contents**

[**1.1 Detailed Background Description** 2](#_Toc155360654)

[**1.2 Materials and Methods** 5](#_Toc155360655)

[**1.3 Sample preparation** 6](#_Toc155360656)

[**1.4 Liquid Chromatograph-Tandem Mass Spectrometry** 7](#_Toc155360657)

[**1.5 Applications** 9](#_Toc155360658)

## **1.1 Detailed Background Description**

Traditional Chinese medicines (TCMs) have been used in primary healthcare for centuries, especially in Europe and Asia. Prescriptions are the main clinical application of TCMs, typically comprising two or more medicinal ingredients prepared using standardized methods to treat specific conditions. Common forms include Chinese patent medicines, Kampo formulas, and hospital decoctions. A new prescription class called “famous classical formulas” was recently developed with vigorous support from the Chinese government, and is expected to boom in the market. Identifying constituent medicinal plants in prescriptions is critical for new drug development and quality control, in order to avoid safety issues from adulteration or substandard ingredients, as seen in the notorious Longdan Xiegan Wan event involving *Aristolochiae Manshuriensis* *Caulis* (Guan Mutong) in place of *Akebiae Caulis* (Mu Tong) in prescriptions[1, 2]. Medicinal plants identification is particularly important for prescriptions prone to replacement of high-value materials with inferior substitutes or deficient amounts, as well as confusion between similar species. Examples include adulteration of *Panax quinquefolium* L. (Xiyangshen) with regular *Panax ginseng* C. A. Mey. (Renshen)[3].

Individual plant identification can be easily achieved using various techniques like morphology, microscopy, genomics, chromatography, and mass spectrometry, with many successful examples. Automated plant species identification has garnered increasing research attention. However, only two tools are publicly accessible: Leafsnap[4] and Pl@ntNet[5]. Unfortunately, neither software focus on medicinal plants or processed materials. Specifically, they only support images, not experimental data. Hence, Chen et al. developed a basic system (http://www.tcmbarcode.cn) for herbal medicine identification using ITS2+*psb*A-*trn*H barcodes, enabling experimental DNA barcode input to search plants[6]. Recently, Beyramysoltan et al. introduced a user-friendly application "DoPP" for rapid species identification of psychoactive plant materials from DART-MS data[7]. However, these platforms and software are not suitable for de-formulating plant prescriptions, a complex system containing multiple medicinal plants.

Formula clarification has always been extremely challenging for multi-ingredient products, especially TCM prescriptions. This is because: 1) Varied dosage forms like tablets, decoction, and pills exist, destroying intrinsic morphological and DNA evidence; 2) Boiling during preparation causing loss of characteristic compounds reported for individual plants; 3) Mutual interference and ion suppression between multiple components; 4) Differing chemical matrices between plants and formulations, requiring tailored analytical methods per Pharmacopeia monographs, hampering efficiency. This necessitates a “one-size-fits-all” analysis approach monitoring all metabolites to concurrently authenticate multiple plants.

Various studies have been devoted to authenticate medicinal plants from complex matrix. Yang’s group developed a three-dimensional characteristic chromatogram to identify Ginseng from 21 various Chinse Patent medicines[8]. Tu’s group proposed a “binary code” concept combined with LC coupled to scheduled multiple reaction monitroring enabling partial plant identification in 6 TCM prescriptions[9]. Our group have applied LC-MS with multivariate statistics to identify multi-source leeches in Tongxinluo capsules[10], donkey-hide gelatin in commercially available samples[11], different part of *Eucommia ulmoides* in 12 commercial products[12], etc. However, these identifications rely on manual diagnostic makers screening, detection and peak confirmation, which are often linked to chromatographic separation, mass spectrometry response and operators’ experience, making the data processing cumbersome. In contrast, automatic data processing owns obvious advantages. Computers rarely produce random errors, so the entire data processing and evaluation process is repeatable. The data processing often takes just minutes, which is more conducive to handling large sample batches, whereas manual processing can take several hours or even days. Most attractively, less prior knowledge from the researcher is required, allowing automatic data processing has been widely accepted. However, a publicly available tool for automated de-formulation of traditional Chinese medicine (TCM) prescriptions is still lacking.

Reported here for the first time are the accomplishment of two aims：1) Developing a universal workflow applicable to all herbal prescriptions; and 2) Creating user-friendly, intuitive software for automated identification of medicinal plants in TCM prescriptions, termed Medicine Check (MedCheck). The application allows users to simply import the LC-dMRM data of an unknown prescriptions into the platform, which then reveals medicinal plant identification with a matching score. It can also be used for the process of raw data to screen diagnostic metabolites to construct in-house database, and for the automatic retention time calibration and confirmation. The performance of the application is demonstrated using homemade and commercial prescriptions, single plants, negatives, and blinded samples.

## **1.2 Materials and Methods**

Nineteen reference standards spanning various compound classes including saponins, flavonoids, alkaloids, iridoids, and lignans were obtained from Shanghai Standard Technology Co., Ltd. (Shanghai, China) and National Institutes for Food and Drug Control (Beijing, China). HPLC-grade acetonitrile (Honeywell, China) and formic acid (TCI Chemical, China), along with analytical-grade methanol (Sinohparm, China) were used. Ultrapure water (18.2 MΩ·cm-1) was redistilled using a Milli-Q Integral water purification system (Merck Millipore, USA). A total of 27 medicinal plants, all in dried form, were collected and identified by Mr. Shuai Yao, a senior medicinal plant identification investigator in our institution, through macroscopic and microscopic characterization. These medicinal plants were used to prepare 7 herbal formulations with varying compositions and proportions.

Chemical profiling was performed on a Waters ACQUITY I-Class UHPLC system coupled to Xevo G2-S Q-TOF mass spectrometer (Waters, Milford, MA, USA) via an ESI interface. Two mobile phases consisting of 0.1% aqueous formic acid (A) and ACN (B) were delivered at a total flow rate of 0.3 mL·min-1. Chromatographic separation was achieved on an HSS T3 column (2.1 mm × 100 mm, 1.7 µm, Waters, USA) maintained at 30 °C using the following gradient elution program: 0-2 min, 5-15% B; 2-8 min, 15-30% B; 8-12 min, 30-65% B; 12-14 min, 65-95% B; 14-16 min, 95-95% B; and 16-20 min, 95-5% B. The injection volume was set at 4.0 µL. High-resolution MS spectra were acquired in Fast DDA mode in the range of m/z 100 to 1200 for MS1 and MS2 in both positive and negative mode with the following parameters: Capillary voltage, 2.5 kV/-2 kV; Sampling Cone, 40 V/-40 V; Source Offset, 80 V/-60 V; Temperature in source, 130 ℃/120 ℃; Temperature for Desolvation, 450 ℃/300 ℃; Gas flow of Cone Gas, 50 L·h-1/50 L·h-1; Gas flow of Desolvation Gas, 800 L·h-1/600 L·h-1; Collision energy ramp, 20-45V and 35-85V.

Method validation and routine testing were performed using an Agilent 1290 UPLC system coupled to an Agilent 6410 triple quadrupole mass spectrometer (TQMS) both in positive and negative ion modes. The chromatographic parameters were identical to Waters ACQUITY I-Class UHPLC system. The electrospray ion source (ESI) parameters were as followings: drying gas (nitrogen) flow rate, 14 L·min-1; drying temperature, 200 ℃; sheath gas flow rate, 11 L·min-1; sheath temperature, 250 ℃; fragment voltage, 135 eV/-130 eV; capillary voltage, 4000 V/-4000 V; collision energy, 35 eV /-30 eV; collision gas, argon. The retention time window in dynamic multi-reaction monitoring (dMRM) mode was 1.0 min. Blank sample for detection of environmental contamination and matrix interference during whole workflow.

Dynamic multi-reaction monitoring, dMRM, is an MS/MS approach based on Multiple Reaction Monitoring (MRM) that is commonly used in MS quantitation. It’s algorithm in LC/MS system automatically constructs retention time window for each analyte throughout the elution time. It allows the instrument to acquire MRM data only during a stated retention time window, thus reducing the number of concurrent ion transitions.

## **1.3 Sample preparation**

Seven homemade TCM prescriptions were prepared according to official protocols, with each containing at least five medicinal plants. The plants were weighed, mixed with water, boiled, and the resulting supernatant freeze-dried into powder. Similarly, 41 negative control samples lacking one specific herb underwent the same preparation process. For comparison, 27 individual medicinal plants were also prepared by weighing, mixing with 10 times the water volume, soaking for 30 minutes, boiling for 45 minutes, and freeze-drying the supernatant to obtain powders. This provided 75 total samples ready for analysis.

For each sample, 0.2 g of extract powder was accurately weighed and soaked in 10 mL of 70% methanol. The mixture was ultrasonically extracted for 1 hour at 37 kHz and 1130 W. After centrifuging at 14,000 rpm for 10 minutes, the supernatant was collected as the test solution.

## **1.4Validation and Applications**

To demonstrate Medcheck's effectiveness, single-blind randomized assays were performed. Thirty blinded samples were randomly selected, including homemade TCM prescriptions and prescriptions lacking specific ingredients (negative controls). Samples were authenticated by matching experimental multiple reaction monitoring (MRM) signals to the diagnostic metabolites in the Medcheck database. Identification results with scores were automatically output. Additionally, 14 commercial TCM prescriptions were obtained: Yinianjin granules, Shiyiwei Shenchen pills, Erchenwan pills, Anshenwan pills, Yiqing granules, Tongxuanlifei granules, Muxiangshunqiwan pills, Danggui Jianzhong Decoction (DGJZD), Banxia Xiexin Decoction (BXXXD), Fuzi Decoction (FZD), Qingweisan Decoction (QWSD), Wenpi Decoction (WPD), Jinshui Liujun Decoction (JSLJD), and Taohong Siwu Decoction (THSWD). Three homemade formulas were also collected by water decocting method: Zhenwu decoction, Shaoyao Gancao decoction, and Lingguizhugan decoction.

## **Reference**

[1] L Wu, B Wang, M M Zhao, W Liu, P Zhang, Y H Shi, et al. Rapid Identification of Officinal Akebiae Caulis and Its Toxic Adulterant Aristolochiae Manshuriensis Caulis *Aristolochia manshuriensis* by Loop -Mediated Isothermal Amplification [J]. Frontiers in Plant Science, 2016, 7.

[2] X Li, Y Xia, G H Li, Z L Zhan, R Y Yao, M H Li. Traditional uses, phytochemistry, pharmacology, and toxicology of Akebiae Caulis and its synonyms: A review [J]. Journal of Ethnopharmacology, 2021, 277.

[3] W Z Yang, X Qiao, K Li, J R Fan, T Bo, D A Guo, et al. Identification and differentiation of *Panax ginseng*, *Panax quinquefolium*, and *Panax notoginseng* by monitoring multiple diagnostic chemical markers [J]. Acta Pharmaceutica Sinica B, 2016, 6(6): 568-75.

[4] W J Kress, C Garcia-Robledo, J V B Soares, D Jacobs, K Wilson, I C Lopez, et al. Citizen Science and Climate Change: Mapping the Range Expansions of Native and Exotic Plants with the Mobile App Leafsnap [J]. Bioscience, 2018, 68(5): 348-58.

[5] M van der Velde, H Goeau, P Bonnet, R d'Andrimont, M Yordanov, A Affouard, et al. Pl@ntNet Crops: merging citizen science observations and structured survey data to improve crop recognition for agri-food-environment applications [J]. Environmental Research Letters, 2023, 18(2).

[6] J P Han, X H Pang, B S Liao, H Yao, J Y Song, S L Chen. An authenticity survey of herbal medicines from markets in China using DNA barcoding [J]. Scientific Reports, 2016, 6.

[7] S Beyramysoltan, M I Chambers, A M Osborne, M I Ventura, R A Musah. Introducing "DoPP": A Graphical User-Friendly Application for the Rapid Species Identification of Psychoactive Plant Materials and Quantification of Psychoactive Small Molecules Using DART-MS Data [J]. Analytical Chemistry, 2022, 94(48): 16570-8.

[8] X Y Xu, M T Jiang, X H Li, Y Wang, M Y Liu, H D Wang, et al. Three-dimensional characteristic chromatogram by online comprehensive two-dimensional liquid chromatography: Application to the identification and differentiation of ginseng from herbal medicines to various Chinese patent medicines [J]. Journal of Chromatography A, 2023, 1700.

[9] Q Q Song, J Li, Y Cao, W J Liu, H X Huo, J B Wan, et al. Binary code, a flexible tool for diagnostic metabolite sequencing of medicinal plants [J]. Analytica Chimica Acta, 2019, 1088: 89-98.

[10] J M Liao, M Gao, Y L Ding, Q R Bi, D D Huang, X X Luo, et al. Characterization of the natural peptidome of four leeches by integrated proteogenomics and pseudotargeted peptidomics [J]. Analytical & Bioanalytical Chemistry, 2023, 415(14): 2795-807.

[11] P L Yang, Q R Bi, Y Li, J M Liao, Y L Ding, D D Huang, et al. Identification of Five Gelatins Based on Marker Peptides from Type I Collagen Mass Spectrum in Multiple Reaction Monitoring Mode [J]. Journal of Agricultural & Food Chemistry, 2023, 71(14): 5851-60.

[12] X Liu, J Q Zhang, Y Li, C L Yao, Y L An, W L Wei, et al. In-depth profiling, nontargeted metabolomic and selective ion monitoring of eight chemical markers for simultaneous identification of different part of *Eucommia ulmoides* in 12 commercial products by UPLC/QDa [J]. Food Chemistry, 2022, 393.