


Article

Utilizing High-Resolution Mass Spectrometry Data Mining Strategy in R Programming Language for Rapid Annotation of Absorbed Prototypes and Metabolites of Gypenosides

Xiaoshan Li ¹, Qianru Zhang ¹, Yuqin Li ¹, Lin Qin ¹, Di Wu ¹, Daopeng Tan ¹, Jian Xie ¹, Jiajia Wu ², Qingping Yang ¹, Yanliu Lu ³, Yongxia Zhao ¹ , Qingjie Fan ¹, Xingdong Wu ^{1,*} and Yuqi He ^{1,*}

- ¹ Guizhou Engineering Research Center of Industrial Key-Technology for Dendrobium Nobile, Zunyi Medical University, Zunyi 563009, China; li_roseixx@163.com (X.L.); zhangqianru@zmu.edu.cn (Q.Z.); lyq1286292@163.com (Y.L.); qinlin1115@163.com (L.Q.); wd_32677@126.com (D.W.); tandp@zmu.edu.cn (D.T.); xiejian@zmu.edu.cn (J.X.); yqp0513@163.com (Q.Y.); x.y.z.100@163.com (Y.Z.); fanqj@zmu.edu.cn (Q.F.)
- ² Shanghai Key Laboratory for Molecular Engineering of Chiral Drugs, School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China; wujiajia_1160@126.com
- ³ Key Lab of the Basic Pharmacology of The Ministry of Education, Zunyi Medical University, Zunyi 563009, China; yanliu.lu@foxmail.com
- * Correspondence: wuxingdong@zmu.edu.cn (X.W.); yqhe.pharm@foxmail.com (Y.H.)

Abstract: The rapid and accurate annotation of the complex compounds and metabolites in natural products remains a significant challenge. In this study, we developed an integrated strategy to efficiently and accurately profile both the prototypes and metabolites of natural products in vivo. And this was achieved by establishing a gypenosides constituent database and utilizing R programming language to combine sample selection, virtual metabolite database construction, polygon mass defect filtering, and Kendrick mass defect filtering. In addition, the annotation strategy was successfully applied to identify the prototypes and metabolites of gypenosides in mice serum. As a case study, gypenoside LXXV was used to validate the feasibility of this approach. The results demonstrated 36 prototypes and 108 metabolites were annotated from the serum by the established annotation strategy. The prototype and eight metabolites of gypenoside LXXV were further confirmed, indicating that the proposed strategy is available. This study provides a novel approach for the rapid and accurate identification of prototypes and metabolites of natural products and offers new insights into the metabolic processes of gypenosides in vivo.

Keywords: R programming language; virtual metabolite library; mass defect; gypenosides; metabolite annotation



Academic Editors: Laura Orian and Giovanni Ribaudo

Received: 26 December 2024

Revised: 29 January 2025

Accepted: 29 January 2025

Published: 7 February 2025

Citation: Li, X.; Zhang, Q.; Li, Y.; Qin, L.; Wu, D.; Tan, D.; Xie, J.; Wu, J.; Yang, Q.; Lu, Y.; et al. Utilizing High-Resolution Mass Spectrometry Data Mining Strategy in R Programming Language for Rapid Annotation of Absorbed Prototypes and Metabolites of Gypenosides. *Molecules* **2025**, *30*, 779. <https://doi.org/10.3390/molecules30040779>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gynostemma pentaphyllum (Thunb.) Makino (GP), also called “Jiao-Gu-Lan”, belongs to the family Cucurbitaceae and genus *Gynostemma* Bl., and is widely distributed in China, Japan, Korea, Southeast Asia and other regions [1]. In 1406 AD, it was first recorded that GP could be eaten as vegetable, and now it is widely used as a health supplement, in the form of a beverage, biscuit, noodles, face wash and bath oil [2]. GP mainly contains various saponins [3], flavonoids [4], amino acids [5], polysaccharides [6] and other components [7], which exhibit a variety of pharmacological properties, including glucose and lipid metabolism regulatory [8,9], anti-inflammatory [10], antioxidant [11], anti-tumor [12], neuroprotective [13] and anti-anxiety properties [14,15]. Modern research has shown that saponins or sapogenins are the main active components responsible for GP’s pharmacological effects [16]. For example, gypenosides (GPs) have been shown to reduce plasma AST,

ALT and TG levels in mice with non-alcoholic fatty liver disease, and to alleviate hepatic steatosis induced by a high-fat, high-cholesterol diet by downregulating ACC1, PPAR, CD36, APOC3 and MTP levels [17]. However, current research on GPs has mainly focused on their pharmacodynamics and pharmacological mechanisms. Consequently, further in-depth studies are required to ascertain whether additional saponins or sapogenins play a role in the pharmacological effects. Hence, the absorbed components characterization of GPs provides a scientific foundation for the discovery of further active components.

The chemical components in natural products undergo phase I or phase II metabolic reactions under the action of enzymes in the body, but their basic skeleton remains unchanged. Therefore, their metabolites share similar characteristics with the structure and fragmentation of the prototype compounds [18]. However, the complexity of the components in natural products, along with the occurrence of one or multiple metabolic reactions in vivo, often results in low concentrations of both the parent compounds and their metabolites, making systematic characterization challenging using traditional methods [19]. LC-MS has become an effective means for chemical composition analysis in natural products due to its multiple scan modes, high resolution and sensitivity [20,21]. However, with the development of mass spectrometry technology, high-resolution mass spectrometry (HRMS) generates vast amounts of data, most of which correspond to unknown metabolites. This poses a significant challenge in identifying target compounds from the data. Recently, an analytical technique based on high resolution mass spectrometry data, mass defect filter, has been developed. It takes advantage of the fact that analogues with the same parent structure, but different substituent groups do not have a large difference in mass loss compared with the parent compound. The mass defect filter can be employed to eliminate all interfering ions that fall outside the desired range, thus facilitating rapid screening of potential components [22]. In addition, after normalizing the m/z value of CH_2 to 14 (the mass of $^{12}\text{C}^1\text{H}_2$), the researchers found that homologues (compounds differ only in the number of CH_2 groups) have the same mass defect (MD) within 5 ppm [23]. This finding has significant implications for annotate drug metabolites if we treat the prototype component entering the body as the compound skeleton and the metabolite as a homologue of the prototype component [24,25].

Although the aforementioned techniques can be used for the rapid screening and characterization of target metabolites in complex systems, with the advancement of mass spectrometry instruments, the volume of acquired mass spectral data increases exponentially. These data often include instrumental noise, spectra generated by impurities in the solvent, and spectra of non-target components, such as adduct ion signals formed between the target component and the solvent [26]. As a result, filtering target mass spectra by manually editing formulas in Excel becomes exceedingly difficult. In recent years, computer programming languages with high-capacity memory, such as R [27], Python [28] and C [29], have been successfully applied to mass spectrometry data processing. Among them, R-based algorithms allow for offline, rapid batch searching of mass spectrometry data for dozens or even hundreds of compounds, including molecular formula, precise molecular weight, and parent ion and daughter ion information under different experimental environments. This significantly reduces the time spent by researchers in online, one-by-one searches [30].

Previously, our research group proposed a concept of using the R programming language to screen and annotate potential prototype components and metabolites in the mice serum administrated with *Dendrobium nobile* extract. However, challenges remained, particularly regarding sample selection and accurate metabolite identification [31]. In this study, based on the established gypenosides component library, an intelligent strategy was developed by integrating R programming language, sample selection, virtual metabolite

library, polygon mass defect filtering and Kendrick mass defect filtering for the rapid and accurate characterization of metabolites in natural products. Additionally, this annotation strategy was applied to identify the prototypic components and metabolites in mice serum administrated with GPs. Finally, the integrated strategy was validated using gypenoside LXXV (Gyp LXXV) as an example. This approach provides a reference for identifying prototypic components and metabolites of natural products *in vivo*, thus supporting the discovery of more active compounds in GPs.

2. Results and Discussion

2.1. Identification of GPs Prototypes

The mass spectrum data of GPs reference standard and GPs were analyzed by UPLC-Q-TOF-MS/MS. 72 chemical components were identified from GPs, with 6 components confirmed by reference standard (Table S1).

To obtain higher-quality mass spectrometry data of serum samples after GPs administration in mice, this study firstly compared the total ion current chromatograms (TIC) of the serum in positive and negative ion modes. The spectral responses in negative ion mode were significantly better than in positive ion mode, allowing for the detection of more serum exogenous constituents (Figure S1). Further analysis of the serum mass spectrometry data revealed the time points at which more exogenous components could be detected (Figure S2). It was observed that more exogenous components could be detected in the serum samples collected at 10, 30 and 45 min, as well as at 1, 8 and 12 h, particularly at 30 min. This suggests that most of the absorbed prototypic constituents or metabolites of GPs could be detected at these time points. However, only a small number of components were detected at 4 h and 6 h, indicating that the constituents from GPs had been mostly eliminated at these time points. These results are consistent with a pharmacokinetics study of GPs, which suggested that the time to peak of macromolecular and small-molecule GP saponins ranged from 7.79 to 12.72 h, and from 2.35 to 6.67 h, respectively [32]. In addition, compared to 4 h and 6 h, the number of detected components at 8 h and 12 h showed an increasing trend, suggesting that absorbed GPs components may be returned to the liver through enterohepatic circulation. To maximize the identification of prototypes and metabolites, serum samples at 10, 30, 45 min and 1, 8, 12 h were pooled and analyzed by UPLC-Q-TOF-MS/MS. The total ion chromatograms of blank serum and drug-containing serum are shown in Figure S3.

To reduce the large amount of redundant data generated by high resolution mass spectrometry, the polygonal mass defect filtering region was established based on the self-built virtual metabolite library to screen the absorbed prototypic components and metabolites of GPs (Figure 1A). Serum samples were analyzed in data-independent acquisition mode, and raw mass spectrometry data were subjected to deconvolution, peak alignment and peak extraction using SCIEX OS (version 3.1.6.44). As a result, 2156 mass spectrometry data were acquired in negative ion mode. After subtracting the blank background, 719 mass spectrometry data were screened out by polygon mass defect filtering, which could be preliminarily identified as prototypic components and metabolites of GPs for further identification, removing about 66% of interfering ions (Figure 1B,C).

Subsequently, blank serum and GPs were used as negative and positive control, respectively. The prototypical GPs components absorbed into the body were identified by comparing the retention time, exact molecular mass, primary and secondary fragments from the mass spectrometry data with the standards, a self-established chemical compound library, and the previously identified GPs components that appeared in both administered serum and GPs but not in blank serum. The characterization process of representative components was as follows: Compound 1, in negative ion mode, exhibited precursor ions

at m/z 883.50 Da and adduct ion peaks at m/z 929.50 Da for $[M+HCOOH-H]^-$ ion. These precursor ions corresponded to a molecular formula of $C_{46}H_{76}O_{16}$ with an error value of 1.7 ppm. In the MS^2 spectrum, fragment ions were detected at m/z 751.46, m/z 605.40 and m/z 473.36 Da. The fragment ions at m/z 751.46 Da primarily resulted from losses of one molecule of xyl from the precursor ion, while the fragment ions at m/z 605.40 Da were attributed to the precursor ions consecutive losing a xyl and rha. Therefore, an aglycone fragment peak at m/z 473.36 Da was obtained through a series of deglycosylation processes. After comparing retention times and mass fragmentation with reference standards, compound **1** was identified as gylongiposide I. The MS and MS^2 spectra, along with potential fragmentation patterns for gylongiposide I, were shown in Figure 2A. Compound **14** in the MS spectrum exhibited $[M-H]^-$ ion peaks at m/z 897.48 Da and $[M+HCOOH-H]^-$ ion peaks at m/z 943.48 Da. In the MS^2 spectrum, the fragment ion at m/z 765.44 Da was generated by losses of one molecule of xyl from the parent ion. While the fragment ions at m/z 681.36 Da were attributed to the precursor ion losing an epoxy (C_5H_8O) in the C-23 position. The fragment ions at m/z 535.33 and m/z 473.36 Da were derived from the precursor ion by losing one molecule of rha and ara, successively. After comparing retention times and mass fragmentation with reference standards, compound **14** was identified as gypenoside A. The MS and MS^2 spectra, along with potential fragmentation patterns for gypenoside A, were presented in Figure 2B. In conclusion, by using a combination of reference standards and literature mass data, a total of 36 prototype components of GPs were identified in mice serum. The extracting ion chromatograms (XIC) were shown as Figure S4, and the detailed information was provided in Table S2.

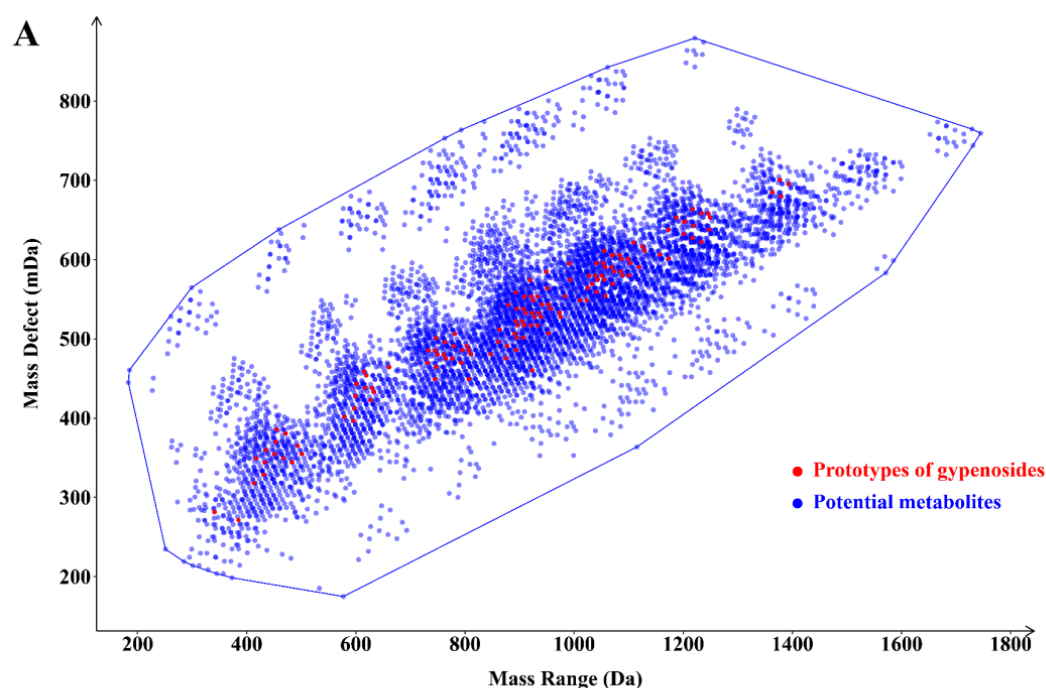


Figure 1. Cont.

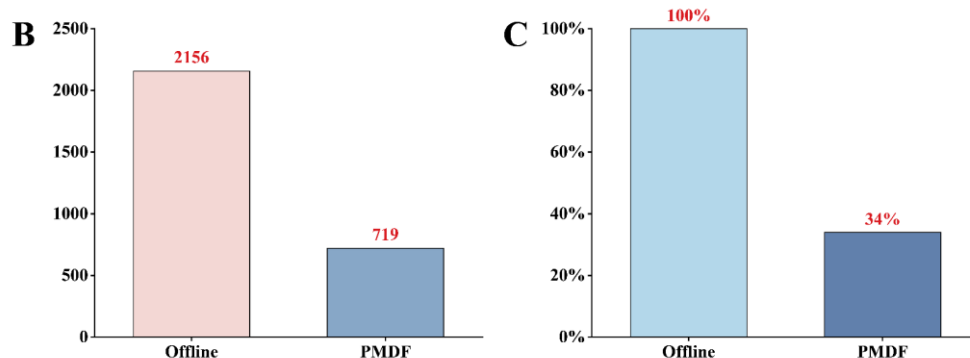


Figure 1. Polygonal mass defect plots for GPs based on the self-built virtual metabolite library and metabolite screening results after administering GPs using the “R” programming language. (A) mass defect regions for potential saponin metabolites in GPs; (B) screening results (absolute value) of potential saponins and their metabolites based on polygonal mass defect in combination with a virtual metabolic product library; (C) screening results (percentage values) of potential saponins and their metabolites based on polygonal mass defect in combination with a virtual metabolic product library.

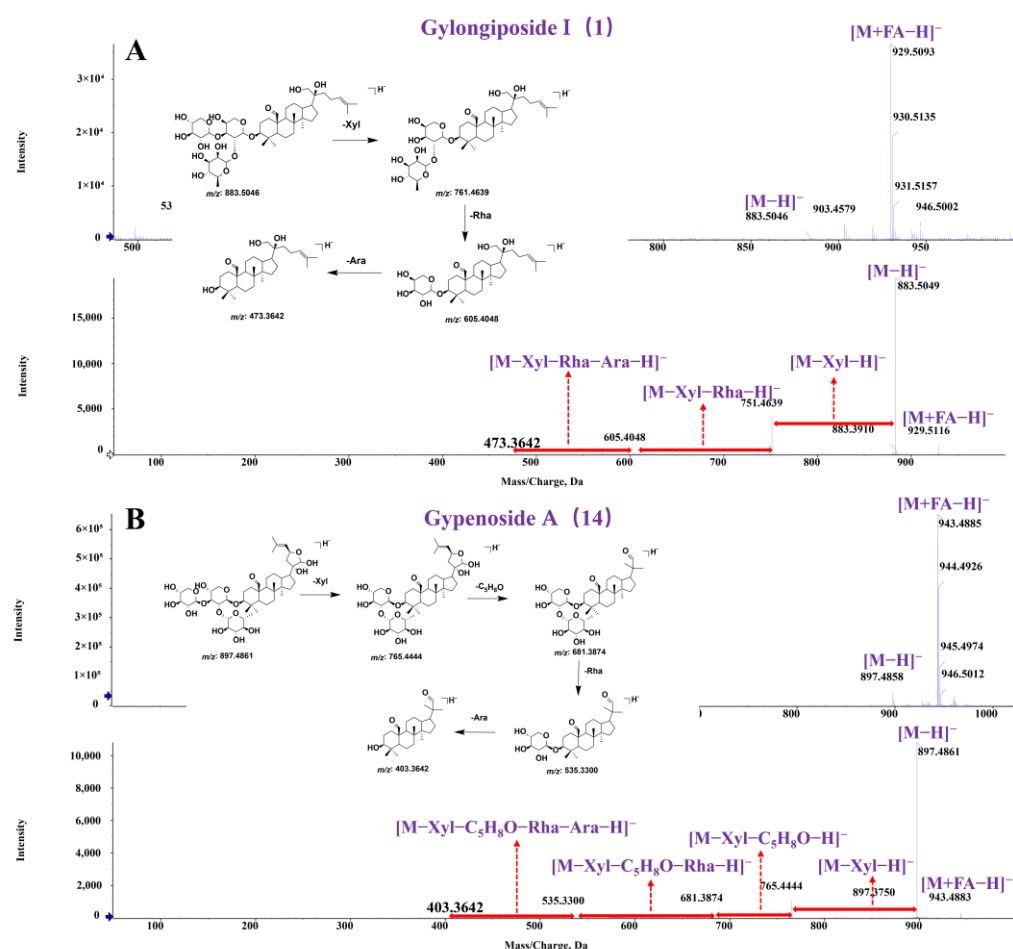


Figure 2. MS and MS² spectra of gypenosides components in mice serum, along with potential mass spectrometry fragmentation patterns. (A) MS and MS² spectra and potential fragmentation patterns of gylongiposide I; (B) MS and MS² spectra and potential fragmentation patterns of gypenoside A.

2.2. Characterization of GPs Metabolites

The chemical composition of natural products is complex and varied, with components present in trace amounts. After entering the body, these components undergo various metabolic transformations in the blood and tissues, resulting in the formation of numerous

metabolites [33,34]. Consequently, the metabolites are found at extremely low concentrations, making their identification challenging. The prototype components and metabolites of GPs in the mouse body were rapidly screened using polygon mass defect and self-built virtual metabolite library. However, the rapid identification of unknown components and metabolic types remains a challenge. Since the mass deficit (MD) theory was first proposed in 1963 [23], the further developed Kendrick mass deficit (KMD) can be used to rapidly characterize homologues in mass spectrometry [35]. All homologues with the same KMD appear as a horizontal line in the plot of nominal Kendrick mass (NKM) (X-axis) against KMD (Y-axis), and the mass-to-charge ratios (m/z) of metabolites within the same composition differ only along the X-axis. Therefore, the method can be used to quickly identify specific metabolite types with known components. In this study, the method was applied to the identification of serum oxygenation and methylation metabolites after the administration of GPs in mice (Figure 3). The peak list data used in Figure 3 is provided in Tables S3 and S4.

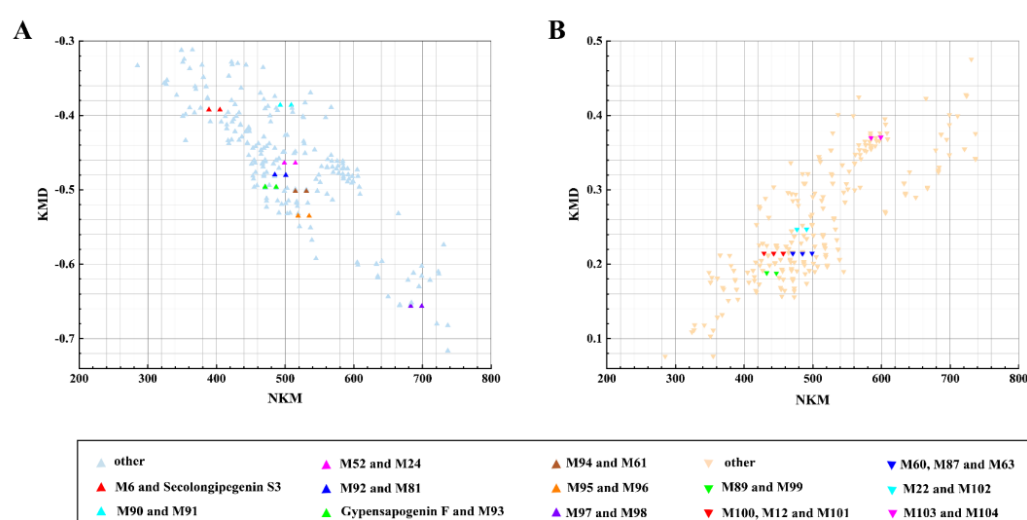


Figure 3. Targeted screening of metabolites of GPs absorbed components in phase I metabolism via deoxidized, oxidation, demethylation and methylation reactions using Kendrick mass defect filtering. (A) Screening process of metabolites after oxidation and deoxidized reactions of GPs; (B) screening process of metabolites after methylation, demethylation reactions of GPs.

Finally, 108 GPs metabolites were quickly screened and characterized (Table S5). Most of these metabolites were phase I metabolites of gypenogenin, including hydroxylation, methylation and alcohol dehydration products. The results of metabolite analysis suggest that GPs may exert their effects through metabolites in the gastrointestinal tract, which is consistent with previous reports [36]. Furthermore, these findings imply that GPs may be absorbed after hydrolysis in the stomach and intestines, followed by metabolic reactions in the body. In addition to the phase I metabolic found in mice serum, phase II metabolites such as acetylation, glucuronic acid binding and sulfate binding of GPs or gypenogenin were also identified. This result aligns with the research conducted by our team, which showed that gypenogenin is metabolized by glucose-aldehyde acidification in human liver microsomal enzymes to produce new products [37]. As an example, the detected gypenoside XLIX (Gyp XLIX) in mice serum was selected for further elucidation of its metabolites and metabolic reaction in vivo (Figure 4). Gylongiposide I was generated by hydrolysis of one glucose at the C-21 position of Gyp XLIX. It is in line with the previous result that Gyp XLIX was transformed into gylongiposide I by the enzymatic biotransformation via the hydrolysis of the glucose group at the C-21 position [38]. Gylongiposide I may be transformed into M54 and M15 by occurring further oxidation and epoxidation,

respectively. In addition, M56 was generated by desaturation on the parent nuclear of gylongiposide I. The metabolite M33 may be obtained by hydroxylation of its methylation product M18 or by further methylation of its hydroxylation product M43.

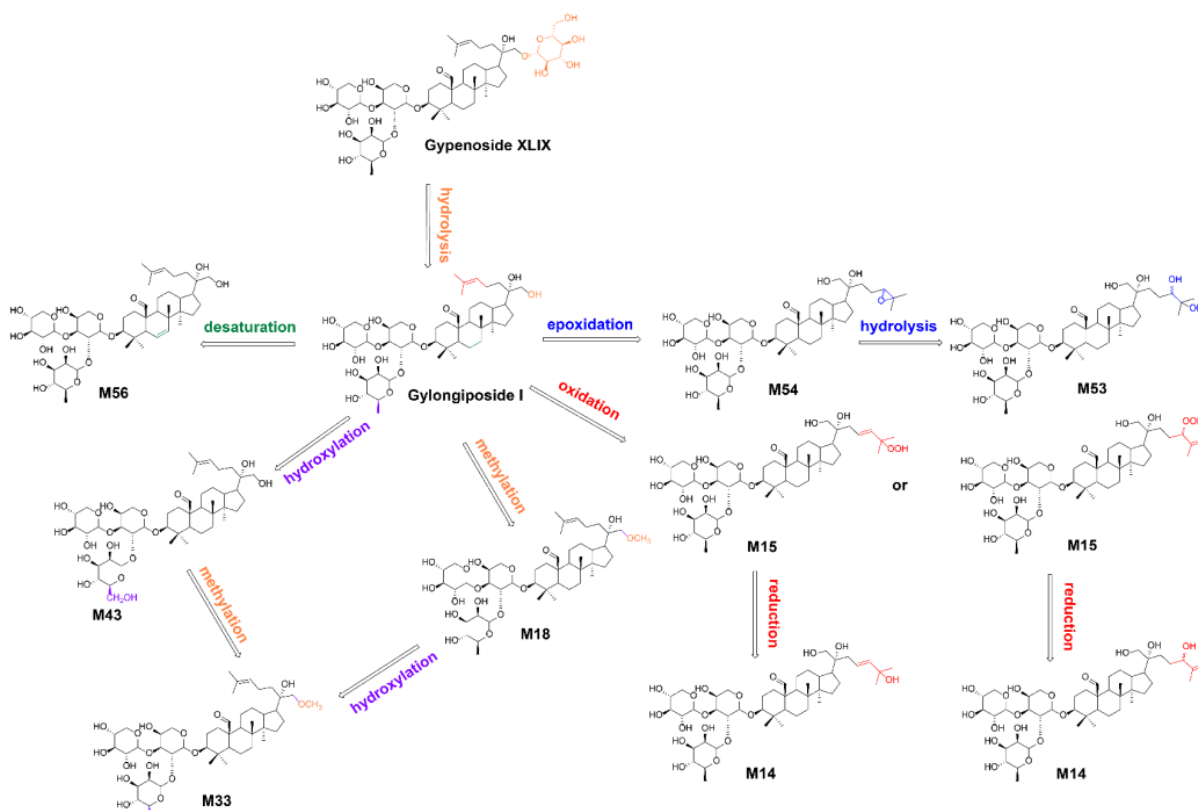


Figure 4. The metabolic pathway of gypenoside XLIX in vivo.

2.3. Application of Rapid Identification Methods for Natural Product Metabolites

Gyp LXXV, a major active component of GPs, was selected as an example to validate the proposed method. To obtain as many metabolites of Gyp LXXV as possible, mice were administered with Gyp LXXV (100 mg/kg) by gavage, and serum mass spectrometry data were collected at 10, 30 and 45 min, and 1, 8 and 12 h. The metabolites were quickly screened and characterized using the integrated strategy to verify the feasibility of the proposed method. As a result, the prototype of Gyp LXXV was detected in mice serum containing Gyp LXXV. Furthermore, this study indicated that Gyp LXXV can be metabolized to Compound K (m/z 621.43 Da) by hydrolyzing and removing a glycosyl group at C-20. In addition, the metabolites of Compound K were further found in both GPs-administrated and Gyp LXXV-administrated serum, including desaturation (M105), oxidation (M108), dealkylation (M107) and glucuronidation (M106) product. However, the metabolites of desaturation and dehydration after dealkylation were found only in the Gyp LXXV-treated serum. This may be due to the mutual interference of various components in the GPs, so that only a small amount of Gyp LXXV is absorbed, resulting in the content of its metabolites being insufficient for further metabolism. According to the metabolites after the administration of Gyp LXXV monomer, the possible metabolic pathways of Gyp LXXV in vivo are shown in Figure 5.

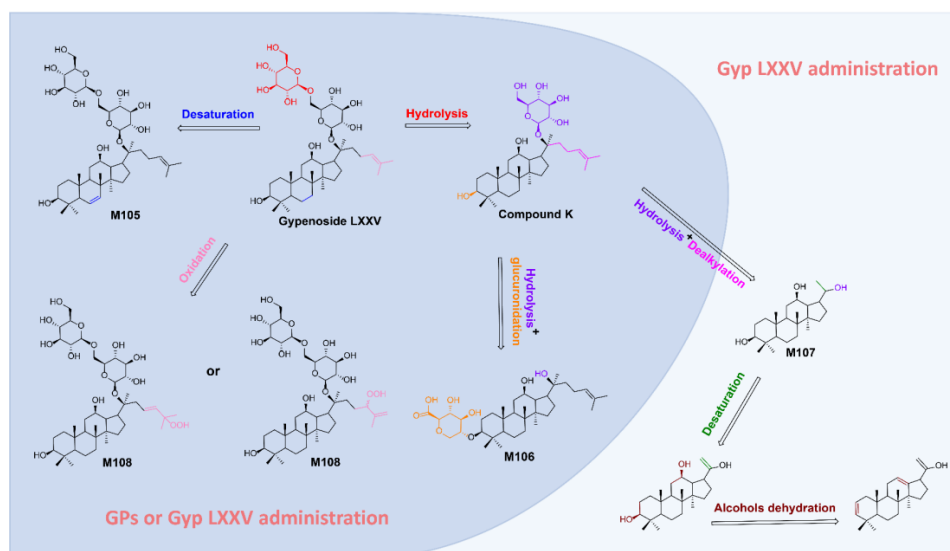


Figure 5. In vivo metabolic pathway of Gyp LXXV in the GPs-administrated and Gyp LXXV-administrated mice.

3. Materials and Methods

3.1. Chemicals and Materials

Gypenosides (purity $\geq 98\%$) were purchased from Shanxi Zhongxin Bio-Technology Co., Ltd., (Yuncheng, China). The reference standards of GPs, Gypenoside A (DST210317-013), Gypenoside XLIX (DST220519-012), Gylongiposide I (DST241023-398), Gypenoside LXXV (DST211020-248), Gypenoside LI (DST211217-280), Gypenoside XVII (DSTD013603), Ginsenoside Rd (DSTD001502) and Gypenoside XLVI (DST190921-111) (purity $\geq 98\%$ by HPLC) were purchased from Chengdu Despite Biotechnology Co., Ltd., (Chengdu, China). The structures of the above eight reference standards are shown in Figure 6. CMC-Na was purchased from Chengdu Kelong Chemical Co., Ltd., (Chengdu, China). Spectro grade methanol, formic acid and acetonitrile were purchased from Thermo Fisher Scientific (Shanghai, China) Co., Ltd. Watson's distilled water was purchased from Watson's food and Beverage Guangzhou Co., Ltd., (Guangzhou, China). All other chemicals were of analytical reagent grade.

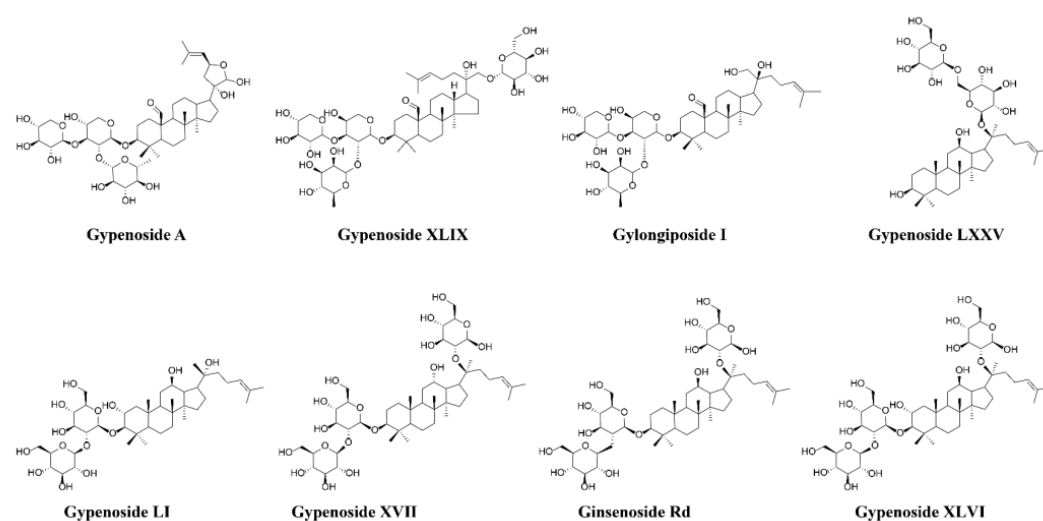


Figure 6. Chemical structures of 8 reference compounds.

3.2. Animals

Thirty healthy male C57BL/6J mal mice (6–8 weeks, weighing 18–22 g) were purchased from Hunan Slack Jingda Experimental Animal Co., Ltd., (Changsha, China, License No. SCXK, 2019-0004). All mice were housed under standard conditions (room temperature 22 ± 1 °C, humidity $55\% \pm 5\%$, 12:12 h light/dark cycle) and were given free access to food and water. Animal care and treatment procedures were approved by the Laboratory Animal Center of Zunyi Medical University (ZMU23-2303-305).

3.2.1. Blood Samples Collection

After acclimating for seven days, all animals were randomly divided into two groups according to body weight. Among them, three mice were assigned to the blank group, and the remaining 27 mice were assigned to the GPs group. The GPs group was orally administered a dose eight times higher than the recommended dose of the 2020 Chinese Pharmacopoeia (216 mg/kg), and the blank group received an equal volume of 0.1% CMC-Na. Blank blood samples were collected from the ophthalmic veins before GPs administration. For the GPs group, blood samples were collected at 10, 30, 45, 60, 120, 240, 360, 480 and 720 min post-administration by anesthetizing three mice at each time point, and blood was collected from the ophthalmic veins (Figure S1). Blood samples were placed at 4 °C for 30 min, centrifuged at 4500 rpm for 15 min, and supernatant was collected. All serum samples were stored at -80 °C until analysis.

3.2.2. Sample Preparation

The serum samples (50 μ L) were mixed with three volumes of acetonitrile (150 μ L), vortexed for 5 min, and centrifuged at 13,000 rpm for 15 min. The supernatant (150 μ L) was evaporated to dryness under nitrogen. The residue was reconstituted in 100 μ L of 100% acetonitrile, vortexed for 5 min and centrifuged. Then, the 50 μ L supernatant was injected into sampling vials for UPLC-Q-TOF-MS/MS analysis.

3.3. UPLC-Q-TOF MS/MS Analysis

A sample analysis was performed on the SHIMADZU LC 40C liquid chromatography system (SHIMADZU, Japan), consisting of an SPD-40V detector, CTO-40C column oven, LC-40B XR solvent delivery unit, SIL-40C autosampler, CBM-40Lite system controller and a Shimadzu LabSolutions workstation. Analytes were separated on an ACQUITY UPLC HSS T3 column (2.1×100 mm, 1.8 μ m). The mobile phase consisted of 0.1% formic acid water (solvent A) and acetonitrile (solvent B). The following optimized UPLC elution conditions were employed: 0% to 15% B at 0–5 min, 15% to 45% B at 5–10 min, 45% to 70% B at 10–20 min, 70% to 100% B at 20–33 min, and finally maintained at 100% B for 3 min. The column temperature, flow rate and injection volume were 30 °C, 0.3 mL/min and 2 μ L, respectively.

Mass spectrometric analysis was performed on an X500R Q-TOF-MS instrument (AB SCIEX, USA) coupled with an electrospray ion source (ESI), and data analysis was conducted using the SCIEX OS (version 3.1.6.44, AB SCIEX, USA) workstation. The operating mass spectrum parameters were as follows: m/z 100 to 1500 Da; ion spray voltage, 4.0 kV (ESI+) and -4.5 kV (ESI $-$); ion source gas 1 (gas 1), 50 psi; ion source gas 2 (gas 2), 50 psi; curtain (CUR) gas, 35 psi; collision associated dissociation (CAD) gas, 8 psi; ion source temperature (Temp), 500 °C. The collision energy (CE) and declustering potential (DP) were set at ± 10 V and ± 50 V, respectively.

3.4. Identification and Characterization of GPs Absorbed Prototypes and Metabolites

3.4.1. Identification of GPs Prototypes

A library for the chemical compound of GPs was established, containing compound names, chemical formula, molecular weight and CAS, sourced from CNKI, PubMed, Web of Science and other databases or platforms (Table S6). Then, the mass spectral information for the identified compounds in the reference, GPs prototype herbs and the gypenosides component database were combined to identify the prototypic components of GPs in mice serum, with a mass spectral deviation controlled within ± 5 ppm.

3.4.2. Characterization of GPs Metabolites

The metabolite identification process was divided into theoretical prediction and manual validation. Theoretically, components entering the body can undergo 71 different metabolic reactions, resulting in the production of 71 metabolites (Table S7). Consequently, the metabolic reactions occurring in vivo for the saponin components of GP were simulated based on the known compounds in GP using the “R” language, resulting in the generation of a virtual metabolite database (Table S8). The code for generating the virtual metabolite database was provided in Supplementary Note S1. To refine the scope of potential metabolites in GPs and acquire higher-quality MS/MS data, this study also utilized “R” language to define the polygonal mass deficit regions for potential metabolites of GPs by using the integer part and decimal part after expanding each metabolite m/z value by a factor of 1000 as the x -axis and y -axis, respectively. Then, the raw data from UPLC-Q-TOF-MS/MS were subjected to deconvolution, peak alignment and peak extraction using SCIEX OS (version 3.1.6.44) to generate a mass spectrometry data matrix for the mice serum administered with GPs. Subsequently, the point. in. polygon function, as part of the SP package within the “R” language, was employed to expeditiously screen the mass spectrometry data within the mass spectrometry data matrix, which conformed to the polygonal mass region for potential prototype components and metabolites of GPs. During the screening process, an allowable range of ± 5 ppm between mass errors was permitted. The code for generating and using polygon mass defect filtering for screening potential prototypes and metabolites of GPs was provided in Supplementary Notes S2 and S3. Next, based on the nominal mass (NMRU) and exact mass (MRU) of the repeating unit (RU), the mass (M) of each screened compound was converted to the Kendrick mass (KM) using the formula (1). Kendrick mass deficit (KMD) was calculated based on the nominal Kendrick mass (NKM) using formula (2) [39]. This process enabled the rapid characterization of methylated, demethylated ($\text{RU} = \text{CH}_2$), oxidative, and deoxidized ($\text{RU} = \text{O}$) metabolites of GPs. The PeakView (version 1.2) was performed for further validation based on precursor ions and metabolite data. The mass spectrum information of the serum sample was compared with the theoretical metabolite data for the 71 predicted metabolites. If the MS and MS^2 data were consistent with the theoretical metabolite data, the metabolite was considered present in the sample. Otherwise, the predicted metabolite does not exist.

$$\text{KM} = \text{M} * \text{NMRU} / \text{MRU} \quad (1)$$

$$\text{KMD} = \text{NKM} - \text{KM} \quad (2)$$

3.5. Application the Feasibility of the Integrated Metabolite Annotation Strategy

To further confirm the feasibility of the proposed strategy and validate the characterization of GPs prototype and metabolite, Gyp LXXV, one of the characterized prototypes, was used as an example to verify its absorption and to further characterize its metabolites. The specific operations conducted were as follows: mice serum samples were collected following the administration of Gyp LXXV (100 mg/kg), and the same chromatographic,

mass spectrometry, and data analysis methods employed in items “2.4” and “2.5” were utilized to identify Gyp LXXV and its metabolites in mice serum.

4. Conclusions

Identifying the prototypic components and metabolites of natural products in vivo remains a challenging task. To quickly and accurately screen and characterize the absorbed prototypes and metabolites of natural products from a large amount of mass spectrometry data, this study developed a rapid and accurate characterization strategy for the prototype components and metabolites of natural products absorbed into the blood by using the R programming language on the basis of the establishment of GPs chemical component database. The proposed strategy was applied to the characterization of prototype components and metabolites in the serum of mice given GPs. As a result, 719 potential metabolites of GPs were screened from mice serum by employing polygonal mass deficit, and 36 prototypic components and 108 metabolites were rapidly characterized in mice serum by employing Kendrick mass deficit. Finally, the strategy was validated using Gyp LXXV as an example, and the results demonstrated that the established method can rapidly identify the metabolites of Gyp LXXV. In addition, most of the metabolites are obtained by phase I metabolic reactions such as hydroxylation and methylation after the hydrolysis of sapogenins, and a few metabolites are phase II metabolites such as acetylation and glucuronidation products of sapogenins. This study not only provides a valuable perspective for the rapid and accurate characterization of prototype components and metabolites of natural products absorbed into the blood but also scientifically elucidates the metabolic pathway of GPs in vivo for the first time.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules30040779/s1>, Figure S1. The total ion current chromatograms (TIC) of blank serum (green) and GPs-administered serum. Figure S2. The proportion of exogenous components detected in mice serum at different time points. Figure S3. The total ion current chromatograms (TIC) of blank serum (red) and GPs-administered serum (green). Figure S4. The extracting ion chromatogram (XIC) of GPs identified absorbed prototypes. Figure S5. The time points of blood samples collection. Table S1. Information on 72 identified chemical components of Gypenosides. Table S2. The 36 prototypes identified from serum administrated with Gypenosides. Table S3. The peak list of screening metabolites after oxidation and deoxidized reactions of GPs. Table S4. The peak list of screening metabolites after methylation, demethylation reactions of GPs. Table S5. The 108 metabolites identified from mice serum administrated with Gypenosides. Table S6. Information on known chemical compounds in Gypenosides. Table S7. Processes of 71 common metabolic reactions in the organism. Table S8. Database of 71 virtual metabolites generated from 333 known chemical compounds of Gypenosides. Supplementary Note S1. R programming code for generating virtual metabolites based on 333 known chemical compounds of Gypenosides. Supplementary Note S2. R programming code for generating PMDF to screen potential absorption prototypes and metabolites of Gypenosides. Supplementary Note S3. R programming code for using PMDF to screen potential absorption prototypes and metabolites of Gypenosides.

Author Contributions: Conceptualization, X.L.; methodology, Q.Z. and Y.L. (Yuqin Li); software, L.Q. and D.W.; writing—original draft preparation, X.L.; validation, D.T. and J.X.; writing—review and editing, X.W.; visualization, J.W. and Q.Y.; supervision, Y.L. (Yanliu Lu), Y.Z. and Q.F.; project administration, Y.H.; funding acquisition, X.W. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the National Nature Science Foundation of China (82260843, 82160740), Guizhou Engineering Research Center of Industrial Key-technology for Dendrobium Nobile (QKJ [2022]048), the Department of Science and Technology of Guizhou Province (QKHPTRC-CXTD [2023]024, QKHPTRC-GCC [2023]075).

Institutional Review Board Statement: The animal study protocol was approved by the Institutional Review Board (or Ethics Committee) of the Laboratory Animal Center of Zunyi Medical University (ZMU23-2303-305).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: We are grateful for all the financial support from the Zunyi Medical University and from the Key Technology Research Center for the Development of the *Dendrobium nobile* Industry, Key Laboratory of Basic Pharmacology, Ministry of Education.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

- Li, Y.; Lin, W.; Huang, J.; Xie, Y.; Ma, W. Anti-cancer effects of *Gynostemma pentaphyllum* (Thunb.) Makino (Jiaogulan). *Chin. Med.* **2016**, *11*, 43. [\[CrossRef\]](#)
- Shi, G.; Wang, X.; Zhang, H.; Zhang, X.; Zhao, Y. New dammarane-type triterpene saponins from *Gynostemma pentaphyllum* and their anti-hepatic fibrosis activities in vitro. *J. Funct. Foods* **2018**, *45*, 10–14. [\[CrossRef\]](#)
- Chen, X.-B.; Yao, C.-L.; Hou, J.-R.; Nie, M.; Li, Y.; Wei, W.-L.; Zhang, J.-Q.; Qu, H.; Li, J.-Y.; Bi, Q.-R.; et al. Systematical characterization of gypenosides in *Gynostemma pentaphyllum* and the chemical composition variation of different origins. *J. Pharm. Biomed. Anal.* **2023**, *232*, 115328. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kao, T.H.; Huang, S.C.; Inbaraj, B.S.; Chen, B. Determination of flavonoids and saponins in *Gynostemma pentaphyllum* (Thunb.) Makino by liquid chromatography–mass spectrometry. *Anal. Chim. Acta* **2008**, *626*, 200–211. [\[CrossRef\]](#)
- Zhang, H.-J.; Gao, P.-F.; Guo, X.-F.; Wang, H. Simultaneous determination of selenium containing amino acids and their sulfur-analogues in green tea and gynostemma pentaphyllum infusion with high performance liquid chromatography based on fluorescence labeling. *Microchem. J.* **2013**, *110*, 192–197. [\[CrossRef\]](#)
- Li, X.-L.; Wang, Z.-H.; Zhao, Y.-X.; Luo, S.-J.; Zhang, D.-W.; Xiao, S.-X.; Peng, Z.-H. Isolation and antitumor activities of acidic polysaccharide from *Gynostemma pentaphyllum* Makino. *Carbohydr. Polym.* **2012**, *89*, 942–947. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nguyen, N.-H.; Ha, T.K.Q.; Yang, J.-L.; Pham, H.T.T.; Oh, W.K. Triterpenoids from the genus *Gynostemma*: Chemistry and pharmacological activities. *J. Ethnopharmacol.* **2021**, *268*, 113574. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xie, P.; Xie, J.-B.; Xiao, M.-Y.; Guo, M.; Qi, Y.-S.; Li, F.-F.; Piao, X.-L. Liver lipidomics analysis reveals the anti-obesity and lipid-lowering effects of gypenosides from heat-processed *Gynostemma pentaphyllum* in high-fat diet fed mice. *Phytomedicine* **2023**, *115*, 154834. [\[CrossRef\]](#)
- Xie, J.B.; Xie, P.; Guo, M.; Li, F.-F.; Xiao, M.-Y.; Qi, Y.-S.; Pei, W.-J.; Luo, H.-T.; Gu, Y.-L.; Piao, X.-L. Protective effect of heat-processed *Gynostemma pentaphyllum* on high fat diet-induced glucose metabolic disorders mice. *Front. Pharmacol.* **2023**, *14*, 1215150. [\[CrossRef\]](#)
- Chen, S.; Liu, X.; Zhao, H.; Cheng, N.; Sun, J.; Cao, W. Beneficial effects of *Gynostemma pentaphyllum* honey paste on obesity via counteracting oxidative stress and inflammation: An exploration of functional food developed from two independent foods rich in saponins and phenolics. *Food Res. Int.* **2022**, *157*, 111483. [\[CrossRef\]](#)
- Ma, C.; Li, H.; Liu, W.; Lu, S.; Li, X.; Chen, J.; Li, K.; Wang, W. Therapeutic Effect of Gypenosides on Antioxidant Stress Injury in Orbital Fibroblasts of Graves' Orbitopathy. *J. Immunol. Res.* **2022**, *2022*, 4432584. [\[CrossRef\]](#)
- Wu, H.; Lai, W.; Wang, Q.; Zhou, Q.; Zhang, R.; Zhao, Y. Gypenoside induces apoptosis by inhibiting the PI3K/AKT/mTOR pathway and enhances T-cell antitumor immunity by inhibiting PD-L1 in gastric cancer. *Front. Pharmacol.* **2024**, *15*, 1243353. [\[CrossRef\]](#) [\[PubMed\]](#)
- Guo, M.; Pei, W.J.; Liu, L.; Chen, K.; Cheng, Y.; Piao, X.-L. Neuroprotective effects of gypenosides on LPS-induced anxiety and depression-like behaviors. *Int. Immunopharmacol.* **2024**, *143*, 113367. [\[CrossRef\]](#)
- Park, H.J.; Zhao, T.T.; Kim, S.H.; Lee, C.K.; Hwang, B.Y.; Lee, K.E.; Lee, M.K. Ethanol extract from *Gynostemma pentaphyllum* ameliorates dopaminergic neuronal cell death in transgenic mice expressing mutant A53T human alpha-synuclein. *Neural Regen. Res.* **2020**, *15*, 361–368.
- Zhao, T.T.; Shin, K.S.; Choi, H.S.; Lee, M.K. Ameliorating effects of gypenosides on chronic stress-induced anxiety disorders in mice. *BMC Complement. Altern. Med.* **2015**, *15*, 323. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xie, P.; Luo, H.-T.; Pei, W.-J.; Xiao, M.-Y.; Li, F.-F.; Gu, Y.-L.; Piao, X.-L. Saponins derived from *Gynostemma pentaphyllum* regulate triglyceride and cholesterol metabolism and the mechanisms: A review. *J. Ethnopharmacol.* **2024**, *319*, 117186. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhou, T.; Cao, L.; Du, Y.; Qin, L.; Lu, Y.; Zhang, Q.; He, Y.; Tan, D. Gypenosides ameliorate high-fat diet-induced nonalcoholic fatty liver disease in mice by regulating lipid metabolism. *PeerJ* **2023**, *11*, e15225. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lin, T.-L.; Lu, C.-C.; Lai, W.-F.; Wu, T.-S.; Lu, J.-J.; Chen, Y.-M.; Tzeng, C.-M.; Liu, H.-T.; Wei, H.; Lai, H.-C. Role of gut microbiota in identification of novel TCM-derived active metabolites. *Protein Cell* **2020**, *12*, 394–410. [\[CrossRef\]](#) [\[PubMed\]](#)

19. Luo, Z.; Xu, W.; Zhang, Y.; Di, L.; Shan, J. A review of saponin intervention in metabolic syndrome suggests further study on intestinal microbiota. *Pharmacol. Res.* **2020**, *160*, 105088. [[CrossRef](#)] [[PubMed](#)]
20. Hou, J.-J.; Zhang, J.-Q.; Yao, C.-L.; Bauer, R.; Khan, I.A.; Wu, W.-Y.; Guo, D.-A. Deeper Chemical Perceptions for Better Traditional Chinese Medicine Standards. *Engineering* **2019**, *5*, 83–97. [[CrossRef](#)]
21. Wu, X.; Hou, J.; Zhang, Z.; Chen, L.; Ni, H.; Qian, Y.; Wu, W.; Long, H.; Zhang, L.; Li, F.; et al. In-depth exploration and comparison of chemical constituents from two *Lilium* species through offline two-dimensional liquid chromatography combined with multimode acquisition of high-resolution mass spectrometry. *J. Chromatogr. A* **2022**, *1670*, 462980. [[CrossRef](#)]
22. Li, D.; Liang, W.; Feng, X.; Ruan, T.; Jiang, G. Recent advances in data-mining techniques for measuring transformation products by high-resolution mass spectrometry. *Trac-Trends Anal. Chem.* **2021**, *143*, 116409. [[CrossRef](#)]
23. Kendrick, E. A Mass Scale Based on CH₂ = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Anal. Chem.* **1963**, *35*, 2146–2154. [[CrossRef](#)]
24. Paulson, A.E.; Lee, Y.J. Novel Ambient Oxidation Trends in Fingerprint Aging Discovered by Kendrick Mass Defect Analysis. *ACS Cent. Sci.* **2022**, *8*, 1328–1335. [[CrossRef](#)]
25. Wu, X.; Liu, Y.; Zhang, Z.; Ou, Z.; Wang, G.; Zhang, T.; Long, H.; Lei, M.; Liu, L.; Huang, W.; et al. Authentication of Shenqi Fuzheng Injection via UPLC-Coupled Ion Mobility—Mass Spectrometry and Chemometrics with Kendrick Mass Defect Filter Data Mining. *Molecules* **2022**, *27*, 4734. [[CrossRef](#)] [[PubMed](#)]
26. Shao, D.; Liu, X.; Wu, J.; Zhang, A.; Bai, Y.; Zhao, P.; Li, J. Identification of the active compounds and functional mechanisms of Jinshui Huanxian formula in pulmonary fibrosis by integrating serum pharmacochimistry with network pharmacology. *Phytomedicine* **2022**, *102*, 154177. [[CrossRef](#)] [[PubMed](#)]
27. Reuschenbach, M.; Hohrenk-Danzouma, L.L.; Schmidt, T.C.; Renner, G. Development of a scoring parameter to characterize data quality of centroids in high-resolution mass spectra. *Anal. Bioanal. Chem.* **2022**, *414*, 6635–6645. [[CrossRef](#)]
28. Volikov, A.; Rukhovich, G.; Perminova, I.V. NOMSpectra: An Open-Source Python Package for Processing High Resolution Mass Spectrometry Data on Natural Organic Matter. *J. Am. Soc. Mass Spectrom.* **2023**, *34*, 1524–1527. [[CrossRef](#)]
29. Duan, L.; Ma, A.; Meng, X.; Shen, G.-A.; Qi, X. QPMAS: A parallel peak alignment and quantification software for the analysis of large-scale gas chromatography-mass spectrometry (GC-MS)-based metabolomics datasets. *J. Chromatogr. A* **2020**, *1620*, 460999. [[CrossRef](#)] [[PubMed](#)]
30. Rosas-Román, I.; Guillén-Alonso, H.; Moreno-Pedraza, A.; Winkler, R. Technical Note: mzML and imzML Libraries for Processing Mass Spectrometry Data with the High-Performance Programming Language Julia. *Anal. Chem.* **2024**, *96*, 3999–4004. [[CrossRef](#)] [[PubMed](#)]
31. Wu, X.; Fan, Q.; Gao, C.; Wu, J.; Wu, D.; Hu, E.; Tan, D.; Zhao, Y.; Li, X.; Zhou, Y.; et al. Metabolites rapid-annotation in mice by comprehensive method of virtual polygons and Kendrick mass loss filtering: A case study of *Dendrobium nobile* Lindl. *J. Pharm. Biomed. Anal.* **2024**, *243*, 116106. [[CrossRef](#)]
32. Zhang, L.; Lin, Y.; Guan, H.; Hu, L.; Pan, G. Simultaneous determination of gypenoside LVI, gypenoside XLVI, 2 α -OH-protopanaxadiol and their two metabolites in rat plasma by LC-MS/MS and its application to pharmacokinetic studies. *J. Chromatogr. B* **2015**, *1005*, 9–16. [[CrossRef](#)] [[PubMed](#)]
33. Yang, M.; Sun, J.; Lu, Z.; Chen, G.; Guan, S.; Liu, X.; Jiang, B.; Ye, M.; Guo, D.-A. Phytochemical analysis of traditional Chinese medicine using liquid chromatography coupled with mass spectrometry. *J. Chromatogr. A* **2009**, *1216*, 2045–2062. [[CrossRef](#)] [[PubMed](#)]
34. Olivon, F.; Apel, C.; Retaillieu, P.; Allard, P.M.; Wolfender, J.L.; Touboul, D.; Roussi, F.; Litaudon, M.; Desrat, S. Searching for original natural products by molecular networking: Detection, isolation and total synthesis of chloroaustralasines. *Org. Chem. Front.* **2018**, *5*, 2171–2178. [[CrossRef](#)]
35. Fouquet, T.; Sato, H. Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units. *Anal. Chem.* **2017**, *89*, 2682–2686. [[CrossRef](#)]
36. Chai, Y.; Luo, J.; Bao, Y. Effects of *Polygonatum sibiricum* saponin on hyperglycemia, gut microbiota composition and metabolic profiles in type 2 diabetes mice. *Biomed. Pharmacother.* **2021**, *143*, 112155. [[CrossRef](#)]
37. Chen, J.; Qin, L.; Wu, X.; Tan, D.; Lu, Y.; Du, Y.; Wu, D.; He, Y. Identification of Human UDP-Glucuronosyltransferase Involved in Gypensapogenin C Glucuronidation and Species Differences. *Int. J. Mol. Sci.* **2023**, *24*, 1454. [[CrossRef](#)]
38. Zhao, H.; Jiao, W.; Xiu, Y.; Zhou, K.; Zhong, P.; Wang, N.; Yu, S. Enzymatic Biotransformation of Gypenoside XLIX into Gylongiposide I and Their Antiviral Roles against Enterovirus 71 In Vitro. *Molecules* **2022**, *27*, 4094. [[CrossRef](#)]
39. Merel, S. Critical assessment of the Kendrick mass defect analysis as an innovative approach to process high resolution mass spectrometry data for environmental applications. *Chemosphere* **2023**, *313*, 137443. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.