ELSEVIER

A targeted strategy to analyze untargeted mass spectral data: Rapid chemical profiling of *Scutellaria baicalensis* using ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry and key ion filtering

CrossMark

Xue Qiao[a], Ru Li[a], Wei Song[a], Wen-juan Miao[a], Jia Liu[b], Hu-biao Chen[c], De-an Guo[a], Min Ye[a,*]

[a] State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China
[b] Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University, 38 Xueyuan Road, Beijing 100191, China
[c] School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong Special Administrative Region, China

ARTICLE INFO

ABSTRACT

Structural identification of natural products by tandem mass spectrometry requires laborious spectral analysis. Herein, we report a targeted post-acquisition data processing strategy, key ion filtering (KIF), to analyze untargeted mass spectral data. This strategy includes four steps: (1) untargeted data acquisition by ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry (UHPLC/orbitrap-MS); (2) construction of a key ion database according to diagnostic MS/MS fragmentations and conservative substructures of natural compounds; (3) high-resolution key ion filtering of the acquired data to recognize substructures; and (4) structural identification of target compounds by analyzing their MS/MS spectra. The herbal medicine Huang-Qin (*Scutellaria baicalensis* Georgi) was used to illustrate this strategy. Its extract was separated within 20 min on a C_{18} column (1.8 μm, 2.1 × 150 mm) eluted with acetonitrile, methanol, and water containing 0.1% formic acid. The compounds were detected in the (−)-ESI mode, and their MS/MS spectra were recorded in the untargeted manner. Key ions were then filtered from the LC/MS data to recognize flavones, flavanones, *O*-/*C*-glycosides, and phenylethanoid glycosides. Finally, a total of 132 compounds were identified from Huang-Qin, and 59 of them were reported for the first time. This study provides an efficient data processing strategy to rapidly profile the chemical constituents of complicated herbal extracts.

1. Introduction

Tandem mass spectrometry (MS/MS) has become an important tool in structural identification of natural products in herbal extracts, which may contain a series of molecules sharing similar scaffolds [1,2]. For unknown compounds, their structural identification mainly follows two routes. The first one is targeted route, where the MS/MS spectra of specific compounds are recorded. This route requires intensive understanding of the analytes and their MS/MS fragmentation pathways prior to analysis, and may sometimes miss important information. The other one is untargeted route, where the MS/MS spectra of all detectable compounds are recorded. Due to the big amount of data obtained, the spectral analysis is usually laborious and time-consuming. Thus far, several methods have been reported for data processing, including fragment classification, spectral similarity evaluation, and informatics tools [3–6]. Herein we report a targeted post-acquisition strategy to analyze untargeted mass spectral data, which could rapidly and accurately profile the chemical constituents of herbal extracts.

This strategy is realized by establishing an ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry method (UHPLC/orbitrap-MS). The Q-Exactive mass spectrometer hybridizes the high mass resolving power (R value up to 140,000) of orbitrap mass analyzer with the selectivity of quadrupole [7]. Multiple precursor ions are fragmented in a high-energy collision cell, and the product ions could be detected with a mass error of less than 3 ppm, even when the analyte concentrations vary in a wide dynamic range [8]. Thus,

Flavone *O*-glycosides

59 R₁=OH, R₂=OgluA, R₃=H	

59 R$_1$=OH, R$_2$=OgluA, R$_3$=H
60 R$_1$=OH, R$_2$=Oglc, R$_3$=H
71 R$_1$=H, R$_2$=OgluA, R$_3$=OH
81 R$_1$=H, R$_2$=OgluA, R$_3$=H
86 R$_1$=OCH$_3$, R$_2$=OgluA, R$_3$=H
87 R$_1$=OCH$_3$, R$_2$=Oglc, R$_3$=H
95 R$_1$=H, R$_2$=OgluA, R$_3$=OCH$_3$

73

17 R$_1$=H, R$_2$=OH, R$_3$=OH
50 R$_1$=OH, R$_2$=H, R$_3$=OCH$_3$

7 R$_1$=R$_4$=OH, R$_2$=R$_3$=H
18 R$_1$=R$_3$=OH, R$_2$=R$_4$=OCH$_3$
23 R$_1$=R$_2$=OCH$_3$, R$_3$=H, R$_4$=OH

Free Flavones

111 R$_1$=OH, R$_2$=OH, R$_3$=H
123 R$_1$=H, R$_2$=OH, R$_3$=OCH$_3$
125 R$_1$=H, R$_2$=OH, R$_3$=H
126 R$_1$=OCH$_3$, R$_2$=OH, R$_3$=H

49 R$_1$=H, R$_2$=OH, R$_3$=OH
108 R$_1$=H, R$_2$=OH, R$_3$=H
126 R$_1$=R$_2$=OCH$_3$, R$_3$=H

Flavone *C*-glycosides

5 R$_1$=glc, R$_2$=xyl, R$_3$=OH
6 R$_1$=xyl, R$_2$=glc, R$_3$=OH
19 R$_1$=α-ara, R$_2$=glc, R$_3$=H
22 R$_1$=β-ara, R$_2$=glc, R$_3$=H
24 R$_1$=glc, R$_2$=α-ara, R$_3$=H
29 R$_1$=glc, R$_2$=β-ara, R$_3$=H
36 R$_1$=H, R$_2$=glc, R$_3$=H

15

Flavanones

9 R$_1$=OH, R$_2$=OH
11 R$_1$=H, R$_2$=Oglc
48 R$_1$=H, R$_1$=OH

94

Phenylethanoid Glycosides

21 R=H
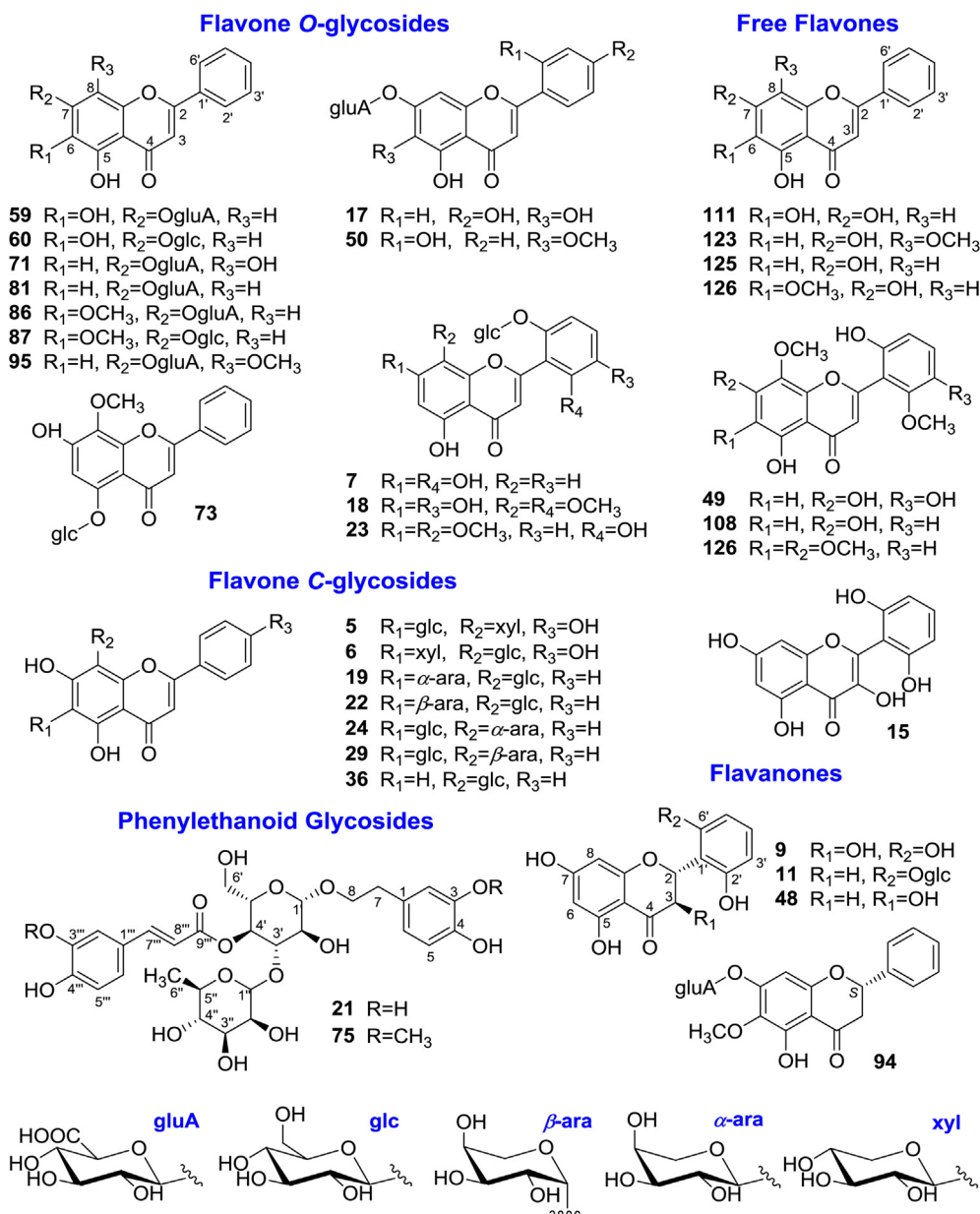75 R=CH$_3$

gluA   glc   β-ara   α-ara   xyl

Fig. 1. Structures of reference standards.
Isoschaftoside (**5**), schaftoside (**6**), 5,7,2′,6′-tetrahydroxyflavone 2′-*O*-β-D-glucoside (**7**), (2*R*,3*R*)-3,5,7,2′,6′-pentahydroxyflavanone (**9**), (2*S*)-5,7,2′,6′-tetrahydroxy flavanone 2′-*O*-β-D-glucoside (**11**), 3,5,7,2′,6′-pentahydroxyflavone (**15**), scutellarin (**17**), viscidulin III 6′-*O*-β-D-glucoside (**18**), chrysin 6-*C*-α-L-arabinopyranoside-8-*C*-β-D-glucoside (**19**), acteoside (**21**), chrysin 6-*C*-β-L-arabinopyranoside-8-*C*-β-D-glucoside (**22**), 5,2′,6′-trihydroxy-7,8-dimethoxyflavone 2′-*O*-β-D-glucoside (**23**), chrysin 6-*C*-β-D-glucoside-8-*C*-α-L-arabinopyranoside (**24**), chrysin 6-*C*-β-D-glucoside-8-*C*-β-L-arabinopyranoside (**29**), chrysin 8-*C*-β-D-glucoside (**38**), (2*S*)-5,7,2′,6′-tetrahydroxyflavanone (**48**), viscidulin III (**49**), 5,7,2′-trihydroxy-6-methoxyflavone 7-*O*-β-D-glucuronide (**50**), baicalin (**59**), baicalein 7-*O*-β-D-glucoside (**60**), norwogonin 7-*O*-β-D-glucuronide (**71**), wogonin 5-*O*-β-D-glucoside (**73**), cistanoside D (**75**), chrysin 7-*O*-β-D-glucuronide (**81**), oroxylin A 7-*O*-β-D-glucuronide (**86**), oroxylin A 7-*O*-β-D-glucoside (**87**), (2*S*)-5,7-dihydroxy-6-methoxyflavanone 7-*O*-β-D-glucuronide (**94**), wogonoside (**95**), 5,7,6′-trihydroxy-8,2′-dimethoxyflavone (**108**), baicalein (**111**), wogonin (**123**), chrysin (**125**), 5,6′-dihydroxy-6,7,8,2′-tetramethoxyflavone (**126**), and oroxylin A (**130**).

high-resolution MS/MS spectral data could be obtained. This technique has rarely been used in natural products analysis [9,10], and no effective strategies have been developed for its data analysis.

Huang-Qin (roots of *Scutellaria baicalensis* Georgi) is one of the most popular herbal medicines worldwide, and is popularly used to treat influenza, pneumonia, and dysentery [11,12]. Due to presence of the predominant compound baicalin (>9%, *w/w*) [11], little attention has been paid to the minor compounds. In the few available reports trying to profile the minor compounds of this plant,

the structures were usually identified by manual interpretation of the tremendous tandem mass spectra [13–15].

In this work, we analyzed the methanol extract of Huang-Qin by UHPLC/orbitrap-MS in the untargeted manner, and the obtained data were processed by a new targeted method called key ion filtering (KIF). A key ion database for Huang-Qin compounds was established, and the acquired data were filtered by the key ions to rapidly identify 132 compounds. UHPLC/orbitrap-MS combined with KIF could be a flexible and powerful strategy to analyze natural products in complex herbal extracts.