Electronic Supplementary material

Computational and data driven-assisted optimizing and understanding of the multiple stage extraction process for polysaccharide and secondary metabolites from natural products

Jiamu Ma[a], Jianling Yao[a], Xueyang Ren[a], Ying Dong[a], Ruolan Song[a], Xiangjian Zhong[a], Yuan Zheng[a], Dongjie Shan [a], Fang Lv [a], Xianxian Li [a], Qingyue Deng [a], Yingyu He [a], Ruijuan Yuan[*a], and Gaimei She[*a]

School of Chinese Meteria Medica, Beijing University of Chinese Medicine, Fangshan District, 100029 Beijing, China

[*]Corresponding author. Tel.: +86-10-5391-2129; Fax: +86-10-5391-2129; E-Mail: rjyuance@126.com; shegaimei@126.com.

Table of Contents

1.General

1.1 List of Figures

1.2 List of Tables

2. Materials and methods

2.1 Measurement and data processing method of panoramic characterization parameters

As the WE process is much more commonly applied in TCM extraction, as well as there are four kinds of materials that take part in the WE process when there is only one material in the EE process, a set of panoramic characterization parameters (qNMR, HPLC fingerprint, and molecular weight of polysaccharide) would explore the changing during the extraction process.

2.2 qNMR analysis procedures

A 2.5 mg BSG was accurately weighed and dissolved in 500 μL of D_2O containing 0.01 mM/mL of DSS-d_6 for the qNMR sample[1,2]. Then the mixed solution was transferred to a 5-mm NMR tube for subsequent analysis. The ^1H NMR was performed by a Bruker Avance 400 MHz NMR spectrometer. The ^1H NMR spectrum acquisition parameters were set as default.

The NMR spectra were processed by Bruker Topspin (Bruker Biospin Corp. Billerica, MA, USA) and MestReNova software (Mestrelab Research, Santiago de Compostela, Spain). The peak of DSS-d_6 was auto-corrected to 0.00 ppm as an internal reference. The embedded modules in Topspin were used to achieve phase correction manually. MestReNova was then used to conduct baseline correction with the Bernstein Polynomial Fit algorithm. The raw spectrum data was handled by binning (δ=0.01 ppm) for dimensionality reduction.

Many peaks have similar chemical shifts (δ, ppm), and overlapped could cause a dilemma like uncertainty quantitative. Therefore, multivariate curve resolution-alternating least squares (MCR-ALS), as a powerful mathematic method, was introduced for recovering the pure NMR spectrum and the corresponding concentration according to the chemical shift[3]. Principal components analysis (PCA), hierarchical cluster analysis (HCA), and partial least squares discriminant analysis (PLS-DA) were applied in the extracted peaks from MCR-ALS for classifying the peaks with a great contribution to the extraction process. For calculating and predicting the extraction process furtherly, the average mass concentration was brought up based on the whole NMR spectrum. Herein, the following equations were presented for describing OED samples:

$$W_x = \frac{A_x}{A_s}\frac{E_x}{E_s}W_s \quad \text{(S1)}$$

Where W_x and W_s are the mass concentrations of components and internal standard, A_x and A_s are the integral areas for characteristic peaks of components and methyl groups of internal standard, and E_x and E_s are the ratios of molar masses to the numbers from the protons, respectively.

$$W_{ave} = W_x \times \sum \frac{W_x}{W_T} \quad \text{(S2)}$$

Where W_{ave} denotes the meaning of average mass concentration, while W_T is the total peak area of all peaks with the chemical shift from 0.5~10.0 ppm.

2.3 HPLC fingerprint analysis

Samples from the WE process of OED were prepared for high-performance liquid chromatography (HPLC) fingerprint analysis. Analyses were performed using a S6000 series HPLC equipment coupled to a PAD detector (Acchorom Tech, China). Compound separation was performed using an Agilent XDB C_{18} column (4.6mm×250mm, 5 μm). The column temperature was set at 30°C. The mobile phases were composed of 0.1 % formic acid in water (solvent A) and acetonitrile (solvent B). The applied gradient was as follows: 0-5 min, 5% A; 5-20 min, 5-10% A; 20-40 min, 10-20% A; 40-55 min, 20-25% A; 55-70 min, 25-30% A; 70-80 min, 30-40% A; 80-85 min, 40-50% A; 85-95 min, 50-60% A; 95-115 min, 60-90% A; 115-120 min, 90%A. The flow velocity and injection volume were 1.0 mL/min and 10 μL. The UV spectra was set at 254 nm due to the maximum number of peaks of samples. The HPLC data files of OED with the suffix of "AIA" were introduced into the Similarity evaluation system for the chromatographic fingerprint of TCM (Version 2012 A) to generate fingerprint chromatographic and identity common peaks. All sets were as default.

Peak area of same peaks as data input to conducted analysis algorithm based on PCA, HCA, and support vector machine (SVM) for clustering. Furthermore, PLS-DA was performed to identify the variable importance for the projection (VIP) peaks and spotting the key components that had a vital influence to sample.

As the clustering results and picked VIP peaks, the average of the area of VIP components was put up as following equation (3) ~ (4). Since the loss of standards and qualitative properties, A_{ave} is taken logarithm as normalization.

$$A_{ave} = A_T \times \sum_{n=1}^{8} \frac{A_n}{A_T} \quad \text{(S3)}$$

$$\overline{A_{ave}} = \log_{10}(A_{ave}) \quad \text{(S4)}$$

Where A_{ave} denotes the average peak area of each OED sample, A_n means the peak area of VIP peaks in each sample, and A_T means the total area of the same peaks in each OED sample. $\overline{A_{ave}}$ means the standardized average peak area of each sample.

2.4 Molecular weight of polysaccharides

Samples to be tested were as same as that in qNMR and HPLC fingerprint. The molecular weight of samples was detected by the HPGPC method on the HPLC (2695, Waters Technology, USA) with an evaporative light-scattering detector (2424, Waters Technology, USA). The sample and standard at a concentration of 5 mg/mL and 1 mg/mL were accurately prepared. TSKgel PWXL 6000 (7.8 mm I.D. × 30 cm, TOSHO, Japan) and TSKgel PWXL 4000 (7.8 mm I.D. × 30 cm, TOSHO, Japan) were equipped for detection. Ultrapure water was used as the mobile phase at a flow rate of 0.5 mL/min. The column temperature was maintained at 30°C. The injection volume was 20 μL. The weight-average (M_W) molecular weight of each peak was determined by retention time (RT) according to the standard curves. The width of each peak, showing the duration of the peak, and the number of peaks was also the features in evaluating the M_W of polysaccharides. To bridge to differences among all OED experiments and to represent the features more specific, average molecular weight (M_{wave}) was adopted. The equation to calculate M_{wave} could be expressed as below:

$$\alpha_n = \frac{A_n}{A_T} \quad (S5)$$

$$M_{wave} = \sum_{n=1}^{4} \alpha_n A_n \quad (S6)$$

$$\overline{M_{wave}} = \log_{10}(M_{wave}) \quad (S7)$$

Where α_n denotes the meaning of the proportion of total peak area accounted for every single peak, A_n is the peak area of a single peak, and A_T refers to the total peak area of all same peaks in OED samples. M_{wave} has the meaning of average molecular weight of polysaccharides in each OED sample, while $\overline{M_{wave}}$ is the standardized form of M_{wave}.

2.5 Content determination of specific parameters in OED

Content of schisandrin and hesperidin were determined by the methods from their original plant recorded by the Chinese Pharmacopoeia 2020 Edition and performed on HPLC (S6000, Acchrom, China). The yield of dry extract was determined under General Provisions 2201 of the Chinese Pharmacopoeia 2020 Edition, either.

Total sugar content was measured using the slightly modified phenol sulfuric acid method. Extract samples to be tested (0.1 g) were homogenized in 10 mL of distilled water. The mixture was dropped into ultrasonic for 30 min. 0.24 mL filtrate was mixed with 1.26 mL 95% ethyl alcohol for precipitating total sugar. While waiting for 8 h, the mixture was centrifuged at 7000 r/min for 20 min. Then, the part of the solid was dissolved in 5 mL water as a sample for the test. The 2 mL sample was mixed with 1.2 mL 5% phenol and 5.5 mL concentrated sulfuric acid, and the reaction was performed under 90℃ water bathing for 15 min. The cool solution was detected for absorbance under 485 nm. All measurements were conducted in triple unless was mentioned specially.

2.6 Determination of components composition of ideal extracts of BWG by UHPLC-Q-Orbitrap-MS/MS analysis

The composition of small molecule compound of BWG was analyzed by an Ultimate 3000 system (Dionex Corp., USA) which was series with Thermo Q Exactive Plus QE (Thermo Fisher Scientific Corp., USA). Data analysis was performed on Compound Discovery 3.2 (Thermo Fisher Scientific Corp., USA). Liquid chromatography analysis was achieved with Agilent ZORBAX Eclipse XDB C_{18} column (250×4.6mm, 5μm). The mobile phases consisted of 0.1% formic acid water (A) and acetonitrile (B) with the gradient elution of 5% B at 0-5 min, 5%-10% B at 5-10 min, 10%-20% B at 10-40 min, 20-25% B at 40-45 min, 25%-30% B at 45-55 min, 30%-45% B at 55-67 min, 45%-50% B at 67-70 min, 50% B at 70-75 min, 50%-60% B at 75-85 min, 60%-75% B at 85-93 min, 75%-90% B at 93-105 min, 90% B at 105-110 min, 90% -5% B at 110-110.1 min, 5% B at 110.1-120 min. The column temperature was 35℃. The flow rate was maintained at 1.0 mL/min, and the injection volume was 5 μL. The acquisition of high-resolution mass spectra was conducted in the positive and negative ion modes. Optimized ionization conditions were as follows: sheath gas flow, 40 L/min; capillary voltage, 3000 V; gas temperature, 320℃; the analysis was carried out using a scan of 75 to 1125 m/z.

2.7 Monosaccharide composition of polysaccharide of the ideal extraction

Monosaccharide composition was determined following PMP derivatization using reverse HPLC[4,5]. In brief, 200 μL 5 mg/mL BWGP solution and 4.0 M TFA solution were mixed, then were hydrolyzed at 110℃ for 120 min. After cooling, 400 μL methanol was added and blown with N_2 to completely remove the excess TFA. The residue was re-dissolved in a mixture of 300 μL solution with isometric of 0.3 M NaOH solution and PMP-methanol solution. Before 100 μL HCl being added, the mixed solution system was at derivatization reaction at 70℃ for 60 min. Then the water was added to 3 mL, and equal volume of chloroform was added either for removing the impurities in the water layer. This process was repeated three times, and filtered with a 0.22 μm membrane as sample to be tested. The sample was analyzed by the same instrumentation and column of HPLC fingerprint analysis. The mobile phase was a mixture of 0.1 M phosphate buffer and acetonitrile (83:17, at pH 6.8), while the flow rate and injection volume were 1.0 mL/min and 10 μL. Standards monosaccharides were processed by using the same method. The content of each monosaccharide in extracts was determined by the calibration curve.

3. Results and discussion

3.1 OED results and the identification of OED extraction process

The simulation conditions based on $L_9(3^4)$ OED of ethanol and water extraction process and also other orthogonal test results are shown in Table S3 – S6. Each row in the table were tested twice, therefore there were 18 sets of raw data for simulating with the help of further analysis. Value of observations in detail is shown in Table S1, Table S2.

Table S1 OED results of the ethanol extraction process of the influencing factors.

Case	Yield of Dry Extraction（%）	Extraction Yield of Schisandrin（%）
1	23.44 ± 0.13	3.81 ± 0.00
2	34.30 ± 0.77	30.35 ± 0.36
3	37.24 ± 0.25	36.05 ± 0.33
4	28.70 ± 0.35	19.55 ± 0.48
5	37.76 ± 0.17	26.84 ± 0.32
6	30.05 ± 0.14	27.68 ± 0.08
7	30.32 ± 0.09	30.88 ± 0.32

8	25.36 ± 0.04	17.53 ± 0.09
9	31.29 ± 0.91	48.84 ± 0.32

Though the importance of view of two evaluating indexes were about to be the same, the slight difference may cause incorrect results of extraction. Taking together, the best ethanol extraction process selected in OED analysis could be summarized as: 2.0 g SC and 24 mL 50% ethanol alcohol was refluxed 2 h, this process repeated twice.

While on the water extraction process, some problems which were same as ethanol extraction process happened again. The water extraction process had more observations, it got more trouble in deciding which extraction process was better. With direct result of OED, the water extraction process was determined as following: 6.0 g AR, 3.0 g BS, 3.0 g CP and the residue of SC after ethanol extraction process was boiled with water for 1.5 hours, this process was repeated twice.

Table S2 OED results of the water extraction process of the influencing factors.

Case	Panoramic Characterization Parameters			Specific Characterization Parameters			
	Wave (μM/mL)	$\overline{A_{ave}}$ (mAu)	$\overline{M_{wave}}$ (kDa)	Yield of Dry Extraction（%）	Content of Total Sugar（mg/g）	Content of Hesperidin（mg/g）	Content of Astragaloside IV（mg/g）
1	0.1192 ± 0.01	5.79 ± 0.01	4.18 ± 0.07	20.39 ± 1.50	8.13 ± 0.16	4.23 ± 0.09	0.3628 ± 0.00
2	0.1703 ± 0.02	5.99 ± 0.02	4.01 ± 0.02	26.54 ± 0.62	11.20 ± 0.32	6.55 ± 0.04	0.5099 ± 0.00
3	0.1085 ± 0.05	6.19 ± 0.04	3.91 ± 0.01	30.05 ± 0.18	11.62 ± 0.21	8.37 ± 0.27	0.5756 ± 0.02
4	0.1364 ± 0.01	6.02 ± 0.05	4.06 ± 0.02	22.30 ± 0.17	8.99 ± 0.08	4.70 ± 0.29	0.5131 ± 0.09
5	0.1945 ± 0.01	5.95 ± 0.03	3.85 ± 0.02	26.82 ± 0.68	9.61 ± 0.23	6.56 ± 0.34	0.5829 ± 0.03
6	0.2205 ± 0.01	6.13 ± 0.07	3.79 ± 0.05	26.68 ± 0.71	10.99 ± 0.21	9.39 ± 0.15	0.6602 ± 0.04
7	0.2209 ± 0.04	5.95 ± 0.01	3.92 ± 0.07	23.62 ± 0.82	5.78 ± 0.02	4.74 ± 0.34	0.4664 ± 0.01
8	0.1696 ± 0.07	5.97 ± 0.06	3.61 ± 0.07	25.77 ± 0.20	7.03 ± 0.04	5.68 ± 0.25	0.6022 ± 0.03
9	0.2851 ± 0.01	6.02 ± 0.02	3.69 ± 0.07	31.93 ± 0.80	10.44 ± 0.09	9.25 ± 0.23	0.5869 ± 0.03

3.2 qNMR analysis for OED

To date, all studies relating to extraction process have utilized chromatographic techniques for quantification of observations like content of some key components and others with more comprehensive characteristic based on retention times. qNMR provides a brand new view for obtaining panoramic condition of sample. Whereas, compared to other spectrum method, qNMR has grown to be a significantly analytical technique of secondary metabolites and the structure of polysaccharide[6,7].

The ^{1}H qNMR spectrum of raw data and dimension by binning data were shown in Fig. 4. Attributed to the dimension of binning processing, the peaks were lightful more concentrate which greatly convenient for the next data analysis. According to the chemical shift of each peak, the spectrum was divided into 3 parts which were high field area (mainly consisting organic acid and some amino acid, ranging 0.50-3.30 ppm), medium field area (mainly was the glycoside components, ranging 3.30-5.50 ppm), and low field area (mainly was the aromatic components, ranging 5.50-10.00 ppm). With the insight of the total area of 3 parts of ^{1}H NMR spectrum, the changing trend of category of components was shown. Along the extraction process, the radio of aromatic components was almost the same, yet the radios between glycoside components, organic acid and amino acid varied a lot. That indicated the extraction process may have noticeable influence on the components like glycoside, organic acid and amino acid, while may affect the components like aromatic little.

To figure out which peaks had close relation with the extraction process, MCR-ALS was introduced for this exploring. 32 peaks were spotted as characteristic peak in the extraction process. Their information in detail was shown in Supplementary material. Then, PCA and HCA results stand the opinion of these peaks were to separate as 3 parts, which mainly suggested their chemical shift was the incontrovertible reason for their classes. PLS-DA acquired 8 peaks as VIP peak among all 32 peaks, also they were noted in Supplementary material. There were 5 peaks coming from the area of medium field area, and 3 peaks could trace to the source of high field area. Glycoside components as well as the small molecule compounds played a vital role of clarifying the extraction process. Herein, the following experiments were to look into the extraction process.

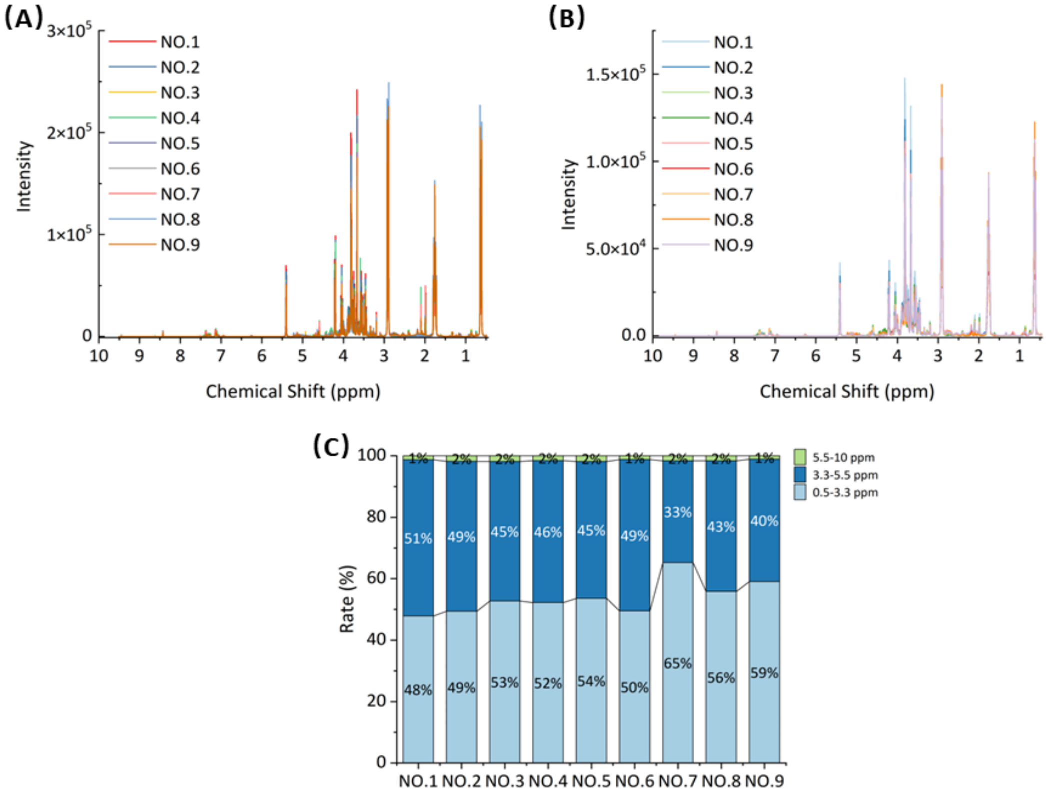

Fig. S 1 Basic information about qNMR analysis. (A) Raw spectrum of 1H-qNMR spectrum, (B) Binning from 0.50-10.00 ppm of 1H-qNMR spectrum, (C) Radio of total peak area in various field.

3.3 HPLC fingerprint analysis for OED

Chromatography fingerprint map is considered to be a reasonable quality evaluation method for TCM and related products[8]. It can capture the slightly difference and has been a powerful methodology to highlight differences during extraction process. Considering about the complex compound composition causing by the formula and their potential interactions, a HPLC fingerprint method was established for analysis the secondary metabolites during extraction process. The HPLC chromatogram of OED samples was shown in Fig. S2. 29 peaks in common (besides the peak of solvent) were gained after basic chromatogram comparison, that were also collected for further analysis. Among which 10 peaks was identified with standards. Hesperidin was set to be the reference peak due to its moderate retention time and suitable area of peak. The HPLC fingerprint was validated by methodological examination of precision, repeatability and stability in 24 h, which had RSD of 0.37%, 2.70% and 0.56%, respectively. The similarity of all samples was within the range of 88.6%~99.5%.

For the better knowing about how the extraction process influence the secondary metabolites, PCA and HCA were conducted for the same peaks in the first place. The area of peak was taken the logarithm for eliminating the huge range of different peak. Peaks were shown gathered into 4 fuzzy groups in PCA, while there were 3 groups in HCA were clearly divided at the same time. SVM is a kind of machine learning algorithm with ability of using a small number of samples for prediction without feature extraction[9]. Therefore, PCA-SVM was introduced for better highlight in clustering the peaks. 95% of classification accuracy was shown in the model of PCA-SVM, while all peaks were divided into 5 parts (Fig. S3). Yet it was still hard for understanding which components or which classification group was more likely impacted by the extraction process with the inefficient cluster in PCA-SVM. The results showed that all peaks could be analyzed by three main components, which accounted for 96.14% of the total variance. PLS-DA was severed for investigating the key peak by VIP value, which 11 peaks in total were characterized. Peak 19, peak 8 (calycosin-7-glucoside), peak 14 (militarine), peak 5 (ferulic acid), peak 3, peak 10 (tangeretin), peak 11 (hesperidin), peak 1, peak 23, peak 9, and peak 6 were selected to be VIP component in analysis. As the retention time from peak 1 to peak 4 were above 4 min, that may be disturbed by various factors and being erratic, therefore they were all get rid of further analysis.

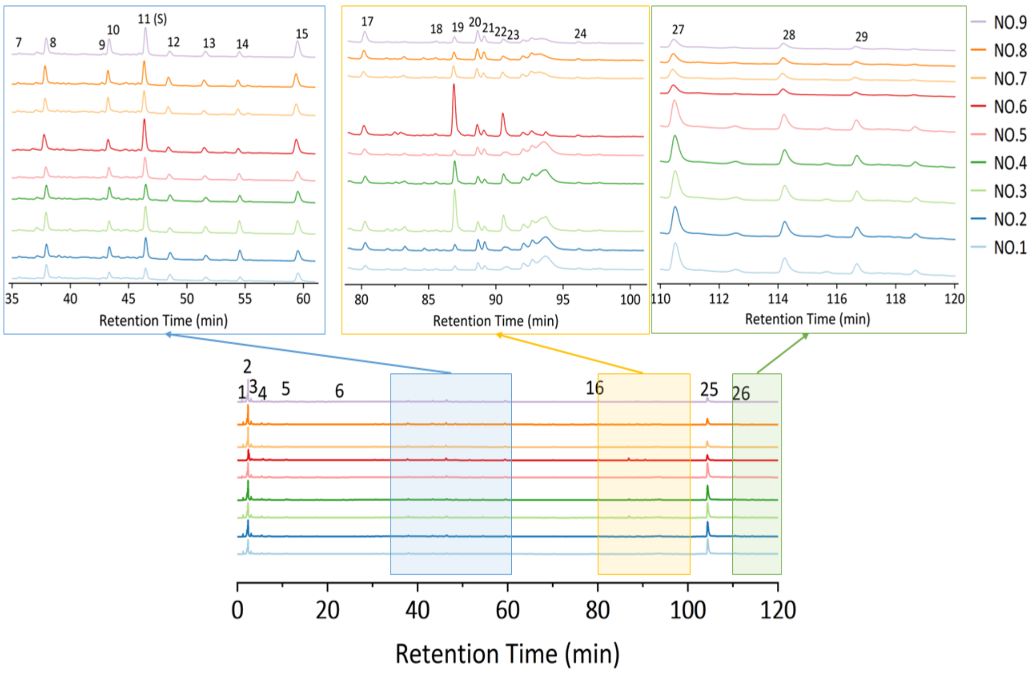

Fig. S 2 HPLC fingerprint schematic diagram of OED. 29 peaks in same were marked.

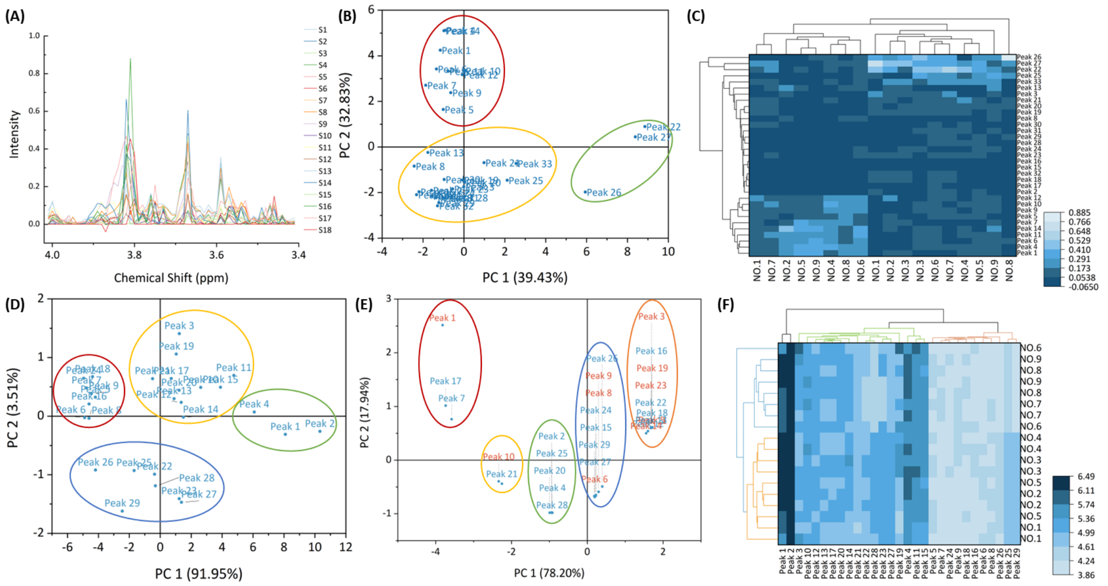

Fig. S 3 Statistical inference of qNMR and HPLC fingerprint based on chemometrics. (A) MCR-ALS resolved spectral profiles of qNMR (3.40-4.00 ppm), (B) PCA of characteristic peak from qNMR, (C) HCA of qNMR, (D) PCA of same peaks from HPLC fingerprint, (E) PCA-SVM of same peaks from HPLC fingerprint based on a Gaussian Kernel, VIP peak was marked as red, (F) HCA of HPLC fingerprint.

3.4 Molecular weight of polysaccharides analysis for OED

M_W is an indispensable factor in the study of the structure-function relationship of polysaccharides, also is an important parameter for emulsifying and rheological properties that influencing digestion and absorption in the body a lot[10–12]. Here, the symbol results of each OED experiment were shown in Fig. S4. It was obvious that extraction influenced a lot in Mw of polysaccharide, the data about M_W of each peak and their duration time in detail was shown in Table.S9. As the figure showing, way of peak being different with other samples could be diverse. There were 5 main peaks in all samples, yet they may be slightly splitting accompany with extraction process. In all, the M_W had trend in being small when the trend of fully extraction performance happened, showing the way was that the retention time of peak was delayed. According to the result shown in Fig.8, BWGP showed heterogeneous structures since there were several elution peaks following elution. As shown in figure, BWGP had 5 peaks, and the mean M_W was 36.35~1301.98 kDa, while peak 1, peak 5 were out of the range of standard linear regression equation. On the other words, M_W of peak 1 was greater than 2457 kDa, M_W of peak 5 were less than 12.6 kDa. Polysaccharide with smaller molecular weights has remarkable bioactivities than the one with larger polysaccharides[13]. It is thought to contribute to smaller polysaccharides pass through biological membranes more easily, and also can more efficiently exert effects without inciting immune stress[14].

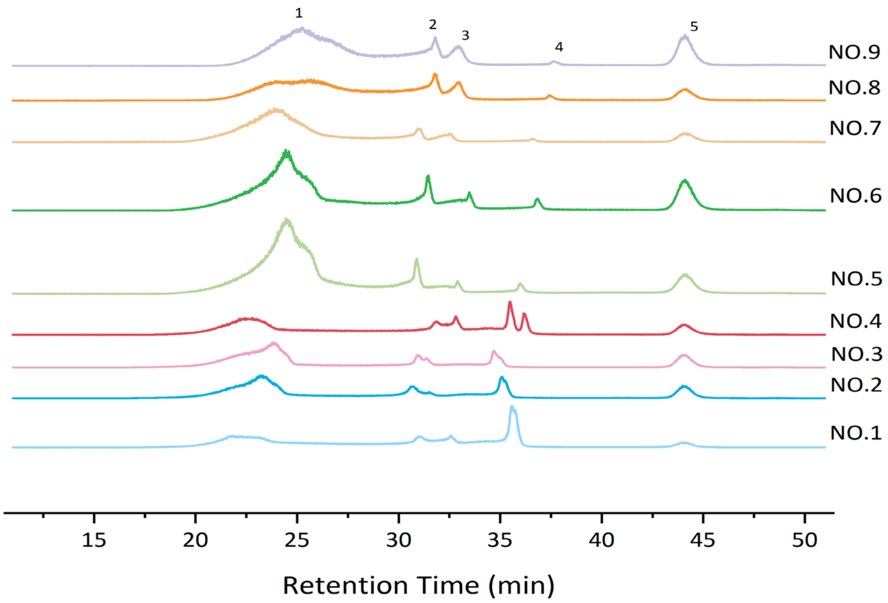

Fig. S 4 Chromatogram of molecular weight.

Table S3. Visual table of OED test in EE process

Experiment number		A	B	C	D	Yield of Dry Extraction（%）	Extraction Yield of Schisandrin（%）
1		1	1	1	1	23.44	3.81
2		1	2	2	2	34.30	30.35
3		1	3	3	3	37.24	36.05
4		2	1	2	3	28.70	19.55
5		2	2	3	1	37.76	26.84
6		2	3	1	2	30.05	27.68
7		3	1	3	2	30.32	30.88
8		3	2	1	3	25.36	17.53
9		3	3	2	1	31.29	48.84
Yield of Dry Extraction	K_1	94.98	82.46	78.85	92.49		
	K_2	96.51	97.42	94.29	94.67		
	K_3	86.97	98.58	105.32	91.3		
	R	9.54	16.12	26.47	3.37		
Extraction Yield of Schisandrin	K_1	70.21	54.24	49.02	79.49		

K_2	74.07	74.72	98.74	88.91
K_3	97.25	112.57	93.77	73.13
R	27.04	58.33	49.72	15.78

Table S4. Analysis of variance table of OED test in EE process

Observation	Source of Variance	Sum of Squared Deviations	Degrees of Freedom	F Radio	F-Critical Value	Significance
Yield of Dry Extraction	A	107.747	2	289.117	19.00	$p<0.01$
	B	235.705	2	632.463	19.00	$p<0.01$
	C	3.882	2	10.417	19.00	$p<0.01$
	D	34.988	2	93.884	19.00	$p<0.01$
Extraction Yield of Schisandrin	A	1167.654	2	6644.884	19.00	$p<0.01$
	B	999.522	2	5688.081	19.00	$p<0.01$
	C	84.106	2	478.631	19.00	$p<0.01$
	D	285.168	2	1622.836	19.00	$p<0.01$

Table S5. Visual table of OED test in WE process

Experiments Number		A	B	C	D	Yield of Dry Extraction (%)	Content of Total Sugar (mg/g)	Content of Hesperidin (mg/g)	Content of Astragaloside IV (mg/g)	W_{ave} (μM/mL)	$\overline{A_{ave}}$ (mAu)	$\overline{M_{wave}}$ (kDa)
1		1:10	1	1	1	20.39	8.125	4.2295	0.3628			
2		1:10	2	1.5	2	26.54	11.195	6.5521	0.5099			
3		1:10	3	2	3	30.05	11.620	8.3655	0.5756			
4		1:12	1	1.5	3	22.30	8.995	4.6957	0.5131			
5		1:12	2	2	1	26.82	9.610	6.5636	0.5829			
6		1:12	3	1	2	26.68	10.99	9.3949	0.6602			
7		1:14	1	2	2	23.62	5.775	4.7400	0.4664			
8		1:14	2	1	3	25.77	7.025	5.6836	0.6022			
9		1:14	3	1.5	1	31.93	10.435	9.2457	0.5869			
Yield of Dry Extraction	K_1	76.98	66.31	72.84	79.14							
	K_2	75.80	79.13	80.77	76.84							
	K_3	81.32	88.66	80.49	78.12							
	R	5.52	22.35	7.93	2.30							
Content of Total Sugar	K_1	30.94	22.895	26.14	28.17							
	K_2	29.595	27.83	30.625	27.96							

	K_3	23.235	33.045	27.005	27.64
	R	7.705	10.15	4.485	0.53
Content of Hesperidin	K_1	19.1471	13.6652	19.308	20.0388
	K_2	20.6542	18.7993	20.4935	20.687
	K_3	19.6693	27.0061	19.6691	18.7448
	R	1.5071	13.3409	1.1855	1.9422
Content of Astragaloside IV	K_1	1.4483	1.3423	1.6252	1.5326
	K_2	1.7562	1.695	1.6099	1.6365
	K_3	1.6555	1.8227	1.6249	1.6909
	R	0.3079	0.4804	0.0153	0.1583
W_{ave}	K_1	0.398	0.4765	0.5093	0.5988
	K_2	0.5514	0.5344	0.5918	0.6117
	K_3	0.6756	0.6141	0.5239	0.4145

	R	0.2776	0.1376	0.0825	0.1972
$\overline{A_{ave}}$	K1	17.97	17.89	17.89	17.76
	K2	18.1	17.91	18.03	18.07
	K3	17.94	18.34	18.09	18.18
	R	0.16	0.45	0.2	0.42
$\overline{M_{wave}}$	K1	12.1	11.58	11.58	11.72
	K2	11.7	11.47	11.76	11.72
	K3	11.22	11.39	11.68	11.58
	R	0.88	0.19	0.18	0.14

Table S6. Analysis of variance table of OED test in WE process

Observations	Source of Variance	Sum of Squared Deviations	Degrees of Freedom	F Radio	F-Critical Value	Significance
Yield of Dry Extraction	A	11.244	2	5.622	19.00	$p<0.01$
	B	167.639	2	83.820	19.00	$p<0.01$
	C	26.979	2	13.489	19.00	$p<0.01$
	D	1.763	2	0.882	19.00	$p>0.05$
Content of Total Sugar	A	22.664	2	11.332	19.00	$p<0.01$
	B	34.216	2	17.108	19.00	$p<0.01$
	C	7.501	2	3.751	19.00	$p<0.01$
	D	0.102	2	0.051	19.00	$p>0.05$
Content of Hesperidin	A	0.781	2	0.390	19.00	$p<0.01$
	B	60.377	2	30.189	19.00	$p<0.01$
	C	0.492	2	0.246	19.00	$p>0.05$
	D	1.304	2	0.652	19.00	$p<0.01$
Content of Astragaloside IV	A	0.034	2	0.017	19.00	$p<0.01$
	B	0.088	2	0.044	19.00	$p<0.01$
	C	6.114×10^{-6}	2	3.057×10^{-6}	19.00	$p>0.05$

Observations	Source of Variance	Sum of Squared Deviations	Degrees of Freedom	F Radio	F-Critical Value	Significance
	D	0.007	2	0.003	19.00	p>0.05
W_{ave}	A	0.026	2	5.514	19.00	p>0.05
	B	0.006	2	0.874	19.00	p>0.05
	C	0.003	2	0.331	19.00	p>0.05
	D	0.016	2	2.727	19.00	p>0.05
$\overline{A_{ave}}$	A	0.142	2	1.054	19.00	p>0.05
	B	0.298	2	2.607	19.00	p>0.05
	C	0.108	2	0.775	19.00	p>0.05
	D	0.126	2	0.921	19.00	p>0.05
$\overline{M_{wave}}$	A	0.254	2	6.644	19.00	p<0.01
	B	0.243	2	6.136	19.00	P<0.05
	C	0.010	2	0.140	19.00	p>0.05
	D	0.010	2	0.137	19.00	p>0.05

Table S7. Same important peaks identification from qNMR in this study.

Chemical Shift (ppm)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
2.93	11454526.91	11537535.28	9266132.33	11345907.25	11939306.88	5761783.11	14164971.04	13224192.55	13839838.27	11302835.55	15038298.25
3.87	975229.42	1050224.14	666964.55	833644.105	966248.245	401955.635	460471.055	994297.705	914281.96	996072.01	959667.195
5.41	5540059.97	5233124.34	3804338.84	4746057.765	4210563.65	3875423.555	3148333.36	4156639.67	4196153.435	3237661.665	3677723.615
0.63	14017726.85	15736797.45	7280167.64	9477185.89	23112447.92	14418125.53	23264025.84	18008448.65	19717640.46	13031761.25	26739496.06
1.79	3135717.735	2831974.69	1517072.445	2519822.53	3124981.665	3749597.96	3775626.97	3228514.295	3449716.14	2543699.16	3542366.54
0.61	9353203.045	10424304.39	9499038.685	10275496	11067877.73	5685545.545	11014359.54	11779540.09	13030474.76	11298351.41	12888226.07
0.64	2920597.54	5785430.57	3827773.26	3891355.81	3730915.68	691056.56	3646904.05	2696641.99	3001857.85	2690787.63	3877442.31

Chemical Shift (ppm)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
	5	5	5	5	5			5		5	5
1.72	2376948.03	2749966.445	2097386.955	2546713.925	2945122.54	1487590.615	3028559.265	3121989.95	3471517.845	2774053.25	3680828.785
1.74	4393430.165	4797865.605	4256608.575	4654720.51	5399286.805	3386934.98	4096960.925	5471307.065	5832777.665	4434929.075	5260206.67
1.75	3658858.08	4447181.405	3038950.925	4055347.53	3995395.755	3451659.425	4748256.595	3921931.61	3577484.79	3910455.08	4412677.69
1.76	10117948.01	10831743.76	9505591.145	10626926.12	12734797.54	7320678.19	12228937.4	13059440.42	14593762.65	11406455.4	13729467.36
1.77	2209420.44	3557639.11	1066403.48	2283943.89	3893544.69	2543846.51	2268515.82	4698722.585	5816705.39	3678646.065	4040190.145
2.90	10292157.63	12025890.66	9430303.525	10870866.82	12484507.14	5769020.595	11440103.47	13465317.66	14373359.97	12303878.05	14530820.74

Chemical Shift (ppm)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
2.91	13705180.14	19991831.26	7650475.755	14739294.57	17357430.02	10667047.36	19282166.62	18675123.35	22781758.61	17903594.05	21752978.66
3.27	610500.68	398881.915	319120.635	439366.71	439388.81	357144.67	424680.51	384021.1	310278.55	343110.145	430678.47
3.34	315945.695	213195.785	194588.37	269700.69	200712.595	249821.5	420467.725	240962.8	353373.595	341220.66	313382.885
3.44	1905065.39	2029251.65	1334418.21	1706544.205	1778605.05	1258737.525	1194303.69	1858770.64	1749658.57	1355073.135	1455821.715
3.49	2329158.52	2577697	1559998.07	2205605.02	2171328.255	1236773.71	1163666.35	1973411.495	2061503.94	1530217.54	1775575.95
3.54	1979526.675	1922769.415	1242027.445	1769980.66	1643855.33	1073117.915	1164809.95	1480813.33	1568927.275	1208677.145	1296559.185
3.56	2590120.53	4588664.065	1854830.895	2989140.255	2725988.33	2216958.655	3329117.55	2735136.945	4734093.71	2747935.8	3093780.545
3.57	2753923.96	3760658.57	2224088.55	1507552.34	2302651.91	2855537.17	2695378.58	1806064.78	3007288.86	1143881.34	1397078.61

Chemi-cal Shift (ppm)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimiza-tion	Computing optimiza-tion
						5	5	5		5	5
3.67	16973641.3	17914732.13	12257128.53	16227635.65	16889342.16	9993111.185	9136190.555	14885578.48	16192876.13	11430701.55	13104142.32
3.78	1979945.705	2293770.17	1388859.535	933069.99	2040419.005	1079684.25	1239000.35	1134147.875	1824931.82	1421835.065	1626468
3.79	165354.835	6834643.79	3989896.29	1002461.77	9681635.635	4665141.34	5054533.975	7645433.425	9125199.525	7008404.88	293365.465
3.80	7705070.51	8092189.185	4455262.92	8005834.345	9827530.83	7554101.14	2886230.065	8999296.99	12373223	5834924.6	7667023.75
3.81	13517552.36	14372303.4	10798712.38	11079150.12	13461077.85	11130638.65	12043723.56	9411795.755	12160475.3	9008590.12	7220166.255
3.88	2237213.165	2713335.05	1464372.285	1891664.75	2596913.865	1527041.14	1301400.04	1969410.85	2031612.03	2161250.7	2021911.61

Chemical Shift (ppm)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
3.89	1064466.255	1644069.135	651470.38	890521.55	850229.09	560913.325	494696.325	1506280.395	856050.085	601645.795	683180.745
3.99	1203077.64	827561.73	524341.915	617671.52	592347.78	403391.975	882815.335	607029.735	457597.465	627471.615	733088.515
4.02	1976254.845	2126845.15	1421364.825	1776329.665	1924774.265	1137067.045	837225.245	1749080.91	1826451.105	1234003.88	1611223.88
4.07	2172445.975	2060923.33	1361528.64	1825813.03	1826427.57	1232368.73	1141126.575	1628125.57	1684746.165	1205091.045	1525086.11
4.20	4058215.82	3875661.26	3066871.175	4252647.925	3359090.24	2235409.98	2039277.82	4182280.69	1063335.305	3227384.055	2320883.11

Table S8. Same peaks identification from HPLC fingerprint analysis in this study.

Retention Time (min)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimiza-tion	Computing optimiza-tion
2.193	661051	1696263.1	1526555.4	1406893	1092076	674929.6	1209560.8	1305836	1616203	1733608	1521460
2.374	2130581	2795863.9	2678988.8	2770879	2007994	1795163	2392183.1	2342362	2533372	3181681	3155345
2.619	175019.2	89110.02	96790.727	43107.71	115807.8	495469.8	85892.764	129392.1	200877.4	84305.35	83074.195
3.046	375220.8	808107.56	578758.47	539068.4	612416.9	419094.5	574043.5	632688.5	822320.7	943303.3	900400.84
9.397	13986.03	17318.34	28303.531	16898.14	18514.82	12232.14	16768.343	18514.77	31811.98	45119.84	36802.572
25.918	18618.83	17521.721	16907.126	19524.32	13058.33	12405.12	14089.803	9164.408	12725.16	12514.25	10746.536

Retention Time (min)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
35.578	15016.88	14993.202	14145.106	11894.19	12447.64	13100.88	18982.963	18772.42	13879.3	19177.75	12904.029
38.950	19826.71	28438.217	21393.042	14578.44	11150.66	14302.22	22881.232	8786.012	12314.76	26396.88	15637.017
42.773	14513.89	16170.941	15775.186	23093.82	11637.36	17294.53	21226.927	17593.44	13150.85	20290.83	18008.108
43.288	95234.65	149137.74	192910.52	214435.4	183270.8	163315	209719.46	195167.9	241496.2	241511.8	201475.31
46.399	192747.4	326629.67	528129.72	275544.1	344142.5	513346.2	392474.81	378426.6	497024.4	343465	425288.8
48.476	117233.1	117590.48	100636.91	120210	95930.91	111619	146504.44	131616.5	103419.7	147096.4	117783.93
51.544	81144.13	106994.25	119950.96	89160.08	98603.23	131598.2	77735.805	95905.4	105521	62004.56	55039.053

Retention Time (min)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
54.464	78940.54	191319.2	209984.63	120213.1	99238.1	86533.75	108925.54	151757.4	106195.6	169350.5	230626.36
59.450	204726.5	263407.88	283359.23	252669.5	237241.5	362701.6	278326.98	257006.1	360258.9	344675.3	452794.3
79.040	15447.84	12535.818	27042.785	20978.44	13604.35	21959.66	12116.338	10673.84	12375.91	12283.37	14009.096
80.223	55446.5	95763.043	116187.73	82028.46	90159.54	122499.3	79846.567	93952.65	138350	126319.2	170404.13
85.526	13668.83	14229.599	14760.402	16468	17886.63	26897.26	15946.327	14911.82	22531.43	16966.27	26654.664
86.906	39384.47	61505.602	376722.99	197315.1	44216.07	462987.9	88296.807	86576.38	73674.33	73253.65	253517.86
88.638	84017.4	115308.79	128183.49	85550.94	109686.1	113128.5	111244	116022.7	141011.2	125843.7	120991.16

Retention Time (min)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
89.126	42243.28	71264.129	49136.66	43164.53	70229.63	69820.87	63701.742	83723.4	85344.78	87425.01	73257.094
92.027	80129.45	98708.531	101139.42	91590.67	92424.27	46298.02	34079.866	41697.25	41817.16	23303.55	28345.684
92.678	158013.3	185791.78	176818.29	176456.4	173640.2	88284.22	36572.336	76490.28	75189.12	25015.44	33639.84
96.132	11081.94	15048.75	15285.811	10148.62	12988.5	14301.43	13642.95	15477.44	18392.83	15611.19	15782.153
103.252	57575.98	59367.954	59234.817	58143.14	57900.19	25975.66	26278.66	27241.22	22915.86	19127.81	19758.927
107.970	22016.04	27644.626	22665.724	29874.05	26537.09	9601.611	8772.7715	12301	14759.94	10911.51	8118.3245
110.466	185020.7	189196.85	189445.29	190120.3	191404.7	58874.67	59669.313	60107.38	57676.96	56161.03	55283.932

Retention Time (min)	Peak area										
	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.7	NO.8	NO.9	OED optimization	Computing optimization
114.193	93810.14	101397.09	100043.12	101176.7	100674	36407.08	38577.194	38430.66	35000.77	33227.17	32459.584
116.653	55064.72	56950.277	55983.512	56479.32	57540.35	14546.37	15386.855	14901.77	13566.44	13705.2	13507.394

Table S9. Same peaks identification from M_W analysis in this study.

Sample	Peak 1		Peak 2		Peak 3		Peak 4		Peak 5	
	M_W (kDa)	Duration Time (min)	M_W (kDa)	Duration Time (min)	M_W (kDa)	Duration Time (min)	M_W (kDa)	Duration Time (min)	M_W (kDa)	Duration Time (min)
NO.1	>2457	17.610-25.070	602.088	30.135-31.810	244.511	31.810-33.130	55.583	37.790-38.550	<12.6	42.520-45.550
NO.2	>2457	16.425-24.676	606.589	29.380-31.100	447.252	31.100-32.250	103.595	34.300-36.500	<12.6	42.54-46.34
NO.3	>2457	17.315-26.150	671.217	29.660-31.220	506.007	31.220-31.970	120.621	33.780-35.950	<12.6	21.130-45.680
NO.4	>2457	17.505-26.623	422.829	30.590-32.320	308.595	32.320-33.650	75.783	35.900-37.000	<12.6	42.250-45.800
NO.5	>2457	18.400-28.100	622.902	29.200-31.700	320.782	32.600-33.600	66.451	35.950-37.250	<12.6	42.300-45.810
NO.6	>2457	18.550-28.300	471.764	29.700-32.200	200.464	32.900-34.050	48.473	36.020-37.600	<12.6	42.200-46.100
NO.7	>2457	18.900-	490.787	29.250-	266.465	31.600-	48.686	35.900-	<12.6	42.200-

		27.900		31.330		34.100		37.700		46.750
NO.8	>2457	19.850-28.900	444.011	30.100-32.100	262.006	32.200-34.700	42.396	36.550-38.175	<12.6	42.250-46.400
NO.9	>2457	20.250-29.200	408.445	30.000-32.350	256.318	32.500-34.000	37.725	36.500-38.350	<12.6	42.500-45.900

Table S10. Components identified in ideal extracts of BSG.

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
Flavonoids and their glycosides								
1	Rutin	$C_{27}H_{30}O_{16}$	19.32	611.1606	611.1620	1.389	303.0564; 153.0194; 285.0760	+H
2	Vicenin II	$C_{27}H_{30}O_{15}$	22.52	593.1500	593.1518	1.764	437.1595; 415.1756; 537.2102; 371.1490	-H
3	Isorhamnetin-3-*O*-neohesperidin	$C_{28}H_{32}O_{16}$	24.33	623.1606	623.1625	0.999	623.1611; 383.0773; 339.0489; 312.0637	-H
4	Narirutin 4'-glucoside	$C_{33}H_{42}O_{19}$	24.65	741.2236	741.2255	1.865	271.0611; 151.0036; 433.1138; 272.0643	-H
5	Naringin-7-*O*-glucoside	$C_{21}H_{22}O_{10}$	24.70	435.1285	435.1293	0.727	200.2374; 201.2409; 116.1434; 57.0701; 71.0779	+H
6	Oroxylin A glucoronide	$C_{22}H_{23}O_{10}$	31.63	447.1285	447.1291	0.547	285.0757; 270.0522; 286.0791	+H
7	Baicalin	$C_{21}H_{18}O_{11}$	31.87	445.0765	445.0803	-2.091	269.0414; 224.0482; 240.0387	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
8	Eriocitrin	$C_{27}H_{32}O_{15}$	31.91	595.1657	595.1674	1.673	151.0035; 135.0451; 287.0561; 107.0138	-H
9	Genistin	$C_{21}H_{20}O_{10}$	32.55	433.1129	433.1131	0.267	433.1496; 271.0591	+H
10	Vicenin III	$C_{26}H_{28}O_{14}$	32.66	563.1395	563.1409	1.458	N/A	-H
11	Quercetin	$C_{15}H_{10}O_{7}$	32.88	303.0499	303.0503	0.391	303.0864; 285.0768; 261.0766; 234.9901; 243.0642; 287.0535	+H
12	Kaempferol 3-gentiobioside	$C_{27}H_{30}O_{16}$	32.96	609.1450	609.1467	1.779	300.0277; 301.0343; 271.0250; 255.0310; 69.0283	-H
13	Apigenin	$C_{15}H_{10}O_{5}$	33.22	269.0444	269.0458	1.39	269.2121; 281.2031	-H
14	Kaempferol 3-neohesperidoside	$C_{27}H_{30}O_{15}$	33.75	593.1500	593.1511	1.084	179.0558; 133.0141	+H
15	Homoplantaginin	$C_{22}H_{22}O_{11}$	36.74	461.1078	461.1095	1.692	417.2987; 373.3093	-H
16	Narirutin	$C_{27}H_{32}O_{14}$	38.51	579.1708	579.1724	1.658	271.0613; 151.0036; 119.0501; 107.0138; 272.0647	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
17	Naringin chalcone	$C_{15}H_{12}O_5$	38.53	273.0757	273.0759	0.210	273.0757; 248.8788; 171.0289; 153.0183	+H
18	7-*O*-Methylluteolin	$C_{16}H_{12}O_6$	39.42	301.0706	301.0707	0.045	286.0485; 153.0504	+H
19	Rhoifolin	$C_{27}H_{30}O_{14}$	39.83	577.1551	577.1567	1.558	269.0456; 68.7911; 99.9255	-H
20	Calycosin-7-*O*-β-D-glucoside	$C_{22}H_{22}O_{10}$	40.39	445.1129	445.1140	1.117	268.0375; 239.0347; 283.0612	-H
21	Rhamnocitrin	$C_{16}H_{12}O_6$	41.50	299.0550	299.0563	1.325	93.0344; 137.0244; 152.8950	-H
22	Diosmin	$C_{28}H_{32}O_{15}$	41.62	609.1813	609.1803	-1.037	303.0863	+H
23	Hesperetin	$C_{16}H_{14}O_6$	43.11	301.0705	301.0719	-0.547	151.0037; 196.0012; 164.0116	-H
24	Neohesperidin	$C_{28}H_{34}O_{15}$	43.11	609.1813	609.1826	1.283	301.0718; 164.0116; 151.0037; 302.0748; 286.0484	-H
25	Hesperidin	$C_{28}H_{34}O_{15}$	43.13	611.197	611.198	1.013	303.0864; 85.0284; 71.0492; 153.0183; 177.0548; 195.0292	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
26	Isosakuranin	$C_{22}H_{24}O_{10}$	43.13	449.1442	449.1449	0.677	195.0290; 263.0551; 85.0285; 245.0445; 219.0291; 177.0547; 165.0185; 69.0336	+H
27	Formononetin	$C_{16}H_{12}O_4$	47.56	267.0651	267.0662	1.035	252.0428; 267.0663; 223.0400; 253.0464; 251.0353	-H
28	Ononin	$C_{22}H_{22}O_9$	47.58	431.1336	431.1342	0.561	269.0808; 241.0864; 137.0232; 133.0645	+H
29	Pinocembrin	$C_{15}H_{12}O_4$	50.00	255.0651	255.0662	1.035	255.2330; 133.7852; 116.9281; 74.0247	-H
30	Tangeretin	$C_{20}H_{20}O_7$	51.02	371.1125	371.1137	1.211	315.2539; 316.2578; 297.2425	-H
31	Eriodictyol	$C_{15}H_{12}O_6$	52.34	283.0550	287.0558	0.805	287.7753; 243.0325	-H
32	(R)-Isomucronulatol	$C_{17}H_{18}O_5$	52.59	303.1227	303.1233	0.62	303.0862; 285.0754	+H
33	Wogonin	$C_{16}H_{12}O_5$	52.71	283.0600	283.0612	1.12	283.0613; 251.0351; 223.0401	-H
34	Calycosin	$C_{16}H_{12}O_5$	52.72	285.0757	285.0759	0.21	285.0758; 270.0522; 137.0234	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
35	Didymin	$C_{28}H_{34}O_{14}$	52.84	593.1864	593.1877	1.258	285.0770; 286.0804; 151.0037	-H
36	Kumatakenin	$C_{17}H_{14}O_6$	54.57	315.0863	315.0865	0.265	269.3028	+H
37	Naringenin	$C_{15}H_{12}O_5$	58.65	273.0759	273.0762	0.511	271.3189; 129.0783; 272.3223; 187.2253	+H
38	Melitidin	$C_{33}H_{40}O_{18}$	59.89	723.2130	723.2149	1.809	57.0345; 402.0956; 387.0727; 359.0774; 417.1190; 329.0305; 99.0450	-H
39	5,7,8,4'-Tetramethoxyflavone	$C_{19}H_{18}O_6$	67.9	343.1176	343.1180	0.395	343.1178; 328.0943; 313.0707; 299.0910	+H
40	6-Demethoxytangeretin	$C_{19}H_{18}O_6$	68.58	341.1019	341.1030	1.125	341.1092; 204.5641; 179.0561; 161.0460	-H
41	Eupatilin	$C_{18}H_{16}O_7$	70.32	343.0812	343.0823	1.141	315.2537; 297.2441; 171.1034	-H
42	Nobiletin	$C_{21}H_{22}O_8$	70.85	403.1387	403.1389	0.236	388.1153; 373.0930; 327.0860; 284.0673	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
43	Isosakuranetin	$C_{16}H_{14}O_5$	71.40	285.0757	285.0769	1.150	285.0769; 164.0114; 196.0013; 286.0800; 136.0167; 151.0036	-H
44	Chrysosplenetin B	$C_{19}H_{18}O_8$	71.50	373.0917	373.0930	1.286	343.0461; 358.0695; 328.0218; 344.0480; 312.9982; 285.0028	-H
45	Isosinensetin	$C_{20}H_{20}O_7$	74.32	373.1281	373.1285	0.391	358.1046; 343.0814; 315.0876; 287.0889	+H
46	Demethylnobiletin	$C_{20}H_{20}O_8$	78.07	389.1230	389.1230	-0.074	359.0780; 341.0656; 285.0761	+H
47	Gardenin B	$C_{19}H_{18}O_7$	83.26	359.1125	359.1128	0.351	N/A	+H
Lignans and their glycosides								
48	Schisanwilsonin A	$C_{27}H_{32}O_9$	56.25	501.2119	501.2123	0.431	479.1825; 418.1969; 375.1808; 83.0492	+H
49	Gomisin S	$C_{23}H_{30}O_7$	68.96	419.2064	419.207	0.66	419.2771; 401.1971; 379.1229; 300.0993	+H
50	Schisandrin	$C_{24}H_{32}O_7$	72.42	433.2220	433.2224	0.762	152.9958; 78.9590	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
51	Schisantherin E	$C_{30}H_{34}O_9$	73.12	539.2275	539.226	-1.489	89.0597; 133.0860; 327.2015; 177.1122; 540.5418	+H
52	Taccalonolide A	$C_{36}H_{46}O_{14}$	73.94	701.2803	701.2821	1.788	717.8382; 236.0506; 115.9207; 99.0452	-H
53	Gomisin J	$C_{22}H_{28}O_6$	75.03	389.1958	389.196	0.215	389.1960; 287.0914; 227.0704; 357.1700	+H
54	Gomisin O/ Gomisin A	$C_{23}H_{28}O_7$	75.90	417.1907	417.1912	0.470	417.2272; 316.1305; 301.1071 347.1490; 285.1118; 418.2307; 402.2038; 242.0937; 317.1349	+H
55	Schisandrin B	$C_{13}H_{10}O$	76.36	183.0785	183.0807	0.318	105.0336; 183.0807; 95.0492; 106.0369	+H
56	Benzoylgomisin H	$C_{30}H_{34}O_8$	79.21	523.2326	523.2312	-1.384	523.2307; 493.1824; 315.1230	+H
57	Schisphenlignan A	$C_{29}H_{30}O_9$	80.42	523.1962	523.2307	-1.934	523.2307; 105.0336; 524.2348	+H
58	Bulbocol	$C_{23}H_{24}O_4$	82.07	365.1747	365.1756	0.864	N/A	+H
59	Schisanhenol	$C_{23}H_{30}O_6$	86.26	403.2115	403.2121	0.615	111.0087; 99.9257; 154.9472; 116.9284	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
60	Gomisin G / Schisantherin A	$C_{30}H_{32}O_9$	87.63	537.2119	537.2101	-1.769	415.1752; 437.1569; 371.1491; 340.1305; 537.2095	+H
61	Schisantherin D	$C_{29}H_{30}O_9$	88.56	521.1806	521.1799	-0.619	184.0735; 104.1070; 86.0964; 124.9999	+H
62	Shancigusin B	$C_{28}H_{26}O_5$	89.43	443.1853	443.1833	-1.98	N/A	+H
63	Schizandrol B	$C_{28}H_{34}O_9$	90.81	512.2275	512.2283	0.831	385.1648; 355.1539; 316.0943; 323.1275; 354.1459; 312.0992;	+H
64	Schizandrin A	$C_{24}H_{32}O_6$	94.57	417.2271	417.2274	0.315	417.2274; 316.1305; 301.1070; 347.1489; 285.1119; 418.2304; 402.2042; 242.0937	+H
65	Schisanlactone B	$C_{30}H_{42}O_4$	95.33	467.3155	467.3162	0.634	95.0855; 449.3051; 431.2940; 1007.0855; 145.1011	+H
66	Schisandrin C	$C_{22}H_{24}O_6$	98.63	385.1645	385.1650	0.445	385.1645; 355.1542; 285.0757; 277.0702	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
67	Schisandrin B	$C_{23}H_{28}O_6$	99.44	401.1958	401.1855	-10.285	401.1961; 402.1996; 386.1726387.1781	+H
68	Schisanlactone D	$C_{30}H_{44}O_3$	102.73	453.3363	453.3363	-0.022	N/A	+H
Other Type								
68	L-Arginine	$C_6H_{14}O_2N_4$	2.30	175.1189	175.1192	0.278	70.0652; 60.0559; 116.0706; 175.1192; 130.0976; 158.0926	+H
69	Betaine	$C_5H_{11}NO_2$	2.49	118.0862	118.0863	0.075	118.0863	+H
70	Sucrose	$C_{12}H_{22}O_{11}$	2.51	341.1078	341.1089	1.082	341.1098; 161.0545	-H
71	Quinic acid	$C_7H_{12}O_6$	2.57	191.0550	191.0562	1.215	N/A	-H
72	Stachydrine	$C_7H_{13}NO_2$	2.62	144.1004	144.1021	0.205	144.1019; 58.0654; 84.0807; 145.1054; 102.2848	+H
73	Adenine	$C_5H_5N_5$	2.71	136.0617	136.062	0.238	136.0217	+H
74	Shikimic acid	$C_7H_{10}O_5$	2.76	173.04444	173.0454	1.040	173.0457; 155.0350	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
75	p-Hydroxybenzoic acid	$C_7H_6O_3$	2.76	137.0233	137.0243	1.029	N/A	-H
76	Niacin	$C_6H_5NO_2$	3.39	124.0393	124.0394	0.155	N/A	+H
77	6-Demethoxytangeretin	$C_6H_8O_7$	3.43	191.0186	191.0199	1.281	N/A	-H
78	Citric acid	$C_6H_8O_7$	3.47	191.0186	191.0198	-0.156	111.0087; 191.0561; 179.0091	-H
79	Catechin	$C_{15}H_{14}O_6$	3.57	289.0706	289.0678	-2.825	290.0885; 128.0353; 133.0143; 115.0036; 200.0564	-H
80	Adenosine	$C_{10}H_{13}N_5O_4$	3.68	268.104	268.1042	0.19	136.0619; 268.1055	+H
81	Gastrodin	$C_{13}H_{18}O_8$	5.25	285.0968	285.0981	1.261	123.0451; 124.0484; 99.9258; 95.0502	-H
82	D (-)-Salicin	$C_{13}H_{18}O_7$	5.25	285.0968	285.0980	1.201	283.8006; 195.0417	-H
83	Vanillic acid	$C_8H_8O_4$	10.11	167.0338	167.0349	1.105	166.9933	-H
84	Protocatechuic acid	$C_7H_6O_4$	10.37	153.0182	153.0193	1.095	N/A	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
85	Parishin E	$C_{19}H_{24}O_{13}$	12.81	459.1133	459.1145	1.183	111.0087; 112.0121; 67.0789	-H
86	4-Hydroxybenzoic acid	$C_7H_6O_3$	14.05	137.0233	137.0244	1.109	N/A	-H
87	Chlorogenic acid	$C_{16}H_{18}O_9$	16.11	355.1023	355.1026	0.331	N/A	+H
88	Caffeic acid	$C_9H_8O_4$	18.42	179.0338	179.0347	0.885	N/A	-H
89	Bletlos B	$C_{27}H_{26}O_7$	19.03	463.1751	463.1711	1.884	463.3031; 127.0367; 377.1919	+H
90	L-Phenylalanine	$C_9H_{11}NO_2$	19.14	166.0862	166.0874	1.145	N/A	-H
91	Ferulic acid	$C_{10}H_{10}O_4$	30.10	193.0495	193.0506	1.105	134.0374; 178.0273; 137.0244	-H
92	1-(p-Hydroxybenzyl)-4,7-dimethoxyphenanthrene-2-ol	$C_{23}H_{20}O_4$	33.58	359.1277	359.1256	-2.116	99.9259; 203.8657; 70.9377; 199.8516; 97.9312	-H
93	Deacetylnomilinic acid	$C_{26}H_{34}O_9$	37.61	491.2275	491.2291	1.631	N/A	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
94	Dactylorhin A	$C_{40}H_{56}O_{22}$	41.31	887.3179	887.3199	2.001	153.0557; 439.1611; 171.0662; 127.0765	-H
95	Emodin	$C_{15}H_{10}O_5$	41.41	269.0444	269.0457	1.300	269.2121; 251.2031; 246.4756	-H
96	Gymnoside II	$C_{21}H_{30}O_{11}$	42.29	457.1704	457.1719	1.472	123.0451; 129.0558; 99.0815; 1217.0764	-H
97	Blestrin C	$C_{30}H_{24}O_6$	42.31	481.1645	481.1691	4.595	487.1440; 415.2113; 384.1931	+H
98	Militarine	$C_{34}H_{46}O_{17}$	49.74	725.2651	725.266	1.474	123.0451; 127.0764; 129.0557; 99.0815; 153.0557; 457.1719	-H
99	3,3’-Dihydroxy-5-methoxy-2,5’,6-tris (p -hydroxybenzyl) bibenzyl	$C_{46}H_{44}O_{11}$	49.80	771.2799	771.272	1.485	123.0451; 127.0764; 153.0559; 129.0558; 99.0813; 171.0658	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
100	3,7-Dihydroxy-2,4-dimethoxyphenanthrene-3-*O*-glucoside	$C_{22}H_{24}O_9$	50.17	431.1337	431.1349	1.261	347.2438; 329.2327; 201.1131	-H
101	2,7-Dihydroxy-4-methoxyphenanthrene-2-*O*-glucoside	$C_{21}H_{22}O_8$	50.27	401.1231	401.1249	1.816	158.9614; 128.9506; 99.0804; 130.9664; 113.9636	-H
102	3,7-Dihydroxy-2,4,8-trimethoxyphenanthrene	$C_{17}H_{16}O_5$	50.83	299.0914	299.0924	1.07	299.2588; 281.2478; 253.2532	-H
103	3'-Hydroxy-5-methoxybibenzyl-3-*O*-β-D-glucopyranoside	$C_{21}H_{26}O_8$	51.59	405.1543	405.1558	1.426	243.1027; 227.0712; 99.9257	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
104	2,7-Dihydroxy-l-(4'-hydroxybenzyl)-9,10-dihydrophenan-threne-4'-*O*-glucoside	$C_{28}H_{30}O_9$	51.75	509.1806	509.1823	1.731	352.8634; 306.7932; 205.8609; 172.8633; 99.9257	-H
105	2,7-Dihydroxy-4-methoxyphenanthrene-2,7-*O*-diglucoside	$C_{27}H_{32}O_{13}$	51.82	563.1759	563.1768	0.843	502.2883; 347.1708; 277.2174	-H
106	Isomucronulatol 7-*O*-glucoside	$C_{23}H_{28}O_{10}$	52.61	461.1598	461.1614	1.567	461.2889; 447.5608; 417.2987; 373.3093	-H
107	2,7-Dihydroxy-3,4-dimethoxyphenanthrene	$C_{16}H_{14}O_4$	55.44	269.0808	268.0823	1.535	252.0428; 267.0663; 223.0400; 253.0464; 251.0353	-H
108	Citrusin III	$C_{36}H_{53}N_7O_9$	57.27	728.3977	728.3982	0.497	662.6917; 419.1340; 389.0870	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
109	2,3,4,7-Tetramethoxyphenanthrene	$C_{18}H_{18}O_4$	57.85	291.1121	291.1134	1.224	291.1608; 250.4938; 141.1286; 121.0294	-H
110	Sitosterol	$C_{19}H_{50}O$	57.99	415.3934	415.3903	-3.093	415.21167	+H
111	Blestritin C	$C_{36}H_{34}O_6$	58.55	563.2428	563.241	-1.785	450.3754; 437.1567; 415.1748	+H
112	2,7-Dihydroxy-1-(p-hydroxybenzoyl)-4-methoxy-9,10-dihydrophene	$C_{22}H_{18}O_5$	61.04	361.107	361.1083	1.29	277.1439; 202.9209	-H
113	Gymnoside V	$C_{49}H_{62}O_{23}$	62.19	1017.3598	1017.362	1.696	N/A	-H
114	Citrusin I	$C_{34}H_{53}N_7O_9$	63.35	702.3821	702.3837	1.627	614.3314; 102.9487; 146.9385; 427.2670; 183.0886; 203.0829	-H
115	Batatasin III	$C_{15}H_{16}O_3$	65.26	245.1172	245.1173	0.169	226.8936; 208.8830; 180.8881	+H
116	Astragaloside I	$C_{41}H_{68}O_{14}$	65.27	783.4525	783.453	0.467	457.6901; 227.1154	-H
117	Astragaloside IV	$C_{41}H_{68}O_{14}$	68.85	785.4681	785.5526	-3.54	455.0798	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
118	3-(4-Hydroxybenzyl)-4-methoxy-2,7-dihydroxy-9,10-dihydrophene	$C_{22}H_{20}O_4$	66.08	349.1434	349.1393	-4.096	259.2458; 231.2146	+H
119	4,7-Dihydroxy-1-(p-hydroxybenzyl)-2-methoxy-9,10-dihydrophene	$C_{22}H_{20}O_4$	66.13	347.1278	347.129	1.214	332.1052; 347.1290; 331.0981; 99.9257	-H
120	Lusianthridin	$C_{15}H_{14}O_3$	66.18	243.1016	243.101	-0.571	242.2842; 243.2877; 208.8836; 158.1905; 57.0702	+H
121	Propindilactone G	$C_{29}H_{38}O_8$	66.74	515.2639	515.2634	-0.515	515.2279; 469.2212; 385.1648; 355.1539	+H
122	1-(4-Hydroxybenzyl)-4-methoxy-2,7-dihydroxyphenanthrene	$C_{22}H_{18}O_4$	67.02	345.1121	345.1134	3.751	345.1133; 330.0897; 302.0954; 237.0557; 99.92577	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
123	1-(p-Hydroxybenzyl)-4,8-dimethoxyphenanthrene-2,7-diol	$C_{23}H_{20}O_5$	68.15	375.1227	375.124	1.380	317.0817; 360.1003; 345.0771; 361.1044; 318.0864; 346.0802	-H
124	Blestrin B	$C_{30}H_{26}O_6$	68.50	483.185	483.1793	-0.865	483.2380; 83.0491; 55.0546;	+H
125	Blestriarene A	$C_{30}H_{26}O_6$	68.55	481.1645	481.1659	1.425	481.1664; 146.9388; 102.9487; 465.1347; 99.9258	-H
126	Limonin	$C_{26}H_{30}O_8$	68.78	469.1856	469.1868	1.166	367.7307; 309.3000; 179.0564; 111.0087	-H
127	3,3'-Dihydroxy-4-(p-hydroxybenzyl)-5-methoxy-bibenzyl	$C_{22}H_{22}O_4$	68.88	349.1434	349.1447	1.274	243.1023; 227.0714; 93.0345; 99.9257; 103.9201	-H
128	3'-*O*-Methylbatatasin III	$C_{15}H_{18}O_3$	68.90	245.1172	245.1175	0.299	226.8936; 208.8831; 180.8882; 224.9272; 197.8909	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
129	Blestrin D	$C_{30}H_{24}O_6$	69.31	479.1489	479.1503	1.415	415.2114; 384.1926; 369.1684	-H
130	Nomilinic acid	$C_{28}H_{36}O_{10}$	69.38	531.2224	531.2239	1.466	59.0138; 99.9259; 60.0170; 325.1830; 414.7184	-H
131	Blestriarene A	$C_{30}H_{22}O_6$	69.84	477.1332	477.135	1.805	293.1260; 107.0492	-H
132	Blestritin B	$C_{30}H_{30}O_6$	70.37	485.1958	485.1958	0.025	146.9386; 123.0452; 102.9487; 99.0815; 127.0765; 129.0559	-H
133	3,3'-Dihydroxy-2,6-bis(p-hydroxybenzyl)-5-methoxy-bibenzyl	$C_{29}H_{28}O_5$	70.74	455.1853	455.1868	1.59	361.1445; 346.1203; 331.0978; 304.1097; 255.1027	-H
134	5,6,7,3',4',5'-Hexamethoxyflavone	$C_{21}H_{22}O_8$	70.85	403.1387	403.1391	0.356	403.1388; 373.0918; 355.0806	+H
135	2,7-Dihydroxy-1,3-di(p-hydroxybenzyl) -4-methoxy-9, 10-dihydrophene	$C_{29}H_{26}O_5$	70.94	453.1696	453.1711	1.46	68.8936; 290.5523; 388.8314; 359.5929; 95.9958; 179.4764	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
136	Bleformin D	$C_{37}H_{32}O_7$	71.57	587.2064	587.2088	2.43	N/A	-H
137	3,3'-Dihydroxy-2-(p-hydroxybenzyl)-5-methoxy-bibenzyl	$C_{22}H_{22}O_4$	72.09	351.159	351.1591	0.064	N/A	+H
138	2,7-Dihydroxy-1-(p-hydroxybenzoyl)-4-methoxy-9, 10-dihydrophene	$C_{22}H_{18}O_5$	72.25	363.1227	363.1235	0.8	N/A	+H
139	(2,3-trans)-2-(4-hydroxy-3-methoxyphenyl)-3-hydroxymethyl-10-methoxy-2,3,4,5-tetrahydro-phenanthro[2,1-b] furan-7-ol	$C_{25}H_{24}O_6$	72.84	419.1489	419.1504	1.565	293.2123; 235.1705; 275.2013; 95.0137	-H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
140	3-Hydroxy-5,6,7,8,3',4'-hexamethoxyflavone	$C_{21}H_{22}O_9$	74.36	419.1336	419.1339	0.281	N/A	+H
141	Blestritin A	$C_{37}H_{36}O_6$	74.66	577.2584	577.2602	1.785	N/A	+H
142	2,6-Bis(4-hydroxybenzyl)-3',5-dimethoxy-3-hydroxyl bibenzyl	$C_{30}H_{30}O_5$	79.02	469.2009	469.2021	1.17	469.3320; 470.3356; 451.3221; 468.3203	-H
143	Pregomisin	$C_{22}H_{30}O_6$	79.51	391.2115	391.2118	0.365	391.2117,167.0704,205.1225; 237.1486; 359.1845	+H
144	4,7,3'5'-Tetramethoxy-9',10'-dihydro-[1,2'-biphenanthrene]-2,7'-diol	$C_{32}H_{27}O_6$	79.56	506.1723	506.1701	-2.29	N/A	-H
145	Bleochranol D	$C_{34}H_{32}O_8$	83.53	569.2169	569.2152	-1.724	431.2065; 387.1803; 356.1612; 68.8067	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
146	Benzoylgomisin Q	$C_{31}H_{36}O_9$	83.63	553.2432	553.2415	-1.629	553.2048; 507.1993; 508.2027; 423.1411; 391.1147	+H
147	Kadsuphilactone B	$C_{30}H_{42}O_5$	84.35	483.3105	483.3106	0.169	483.2379; 427.1748; 414.1678; 395.1487; 451.2117	+H
148	2',6'-Bis(p-hydroxybenzyl)-5-methoxybibenzyl-3,3'-diol	$C_{33}H_{36}O_5$	89.13	511.2479	511.2721	-2.773	103.9202; 104.9279; 71.0138; 184.9544; 158.9386; 114.9487; 187.0062	-H
149	Linolic acid	$C_{16}H_{22}O_4$	92.63	277.1434	277.1446	1.214	277.1446; 121.0294	-H
150	2,7-Dihydroxy-1,3-di(p-hydroxybenzyl)-4-methoxy-9,10-dihydrophene	$C_{29}H_{26}O_5$	94.60	455.1853	455.1832	-2.04	N/A	+H
151	Bleochranol A	$C_{40}H_{38}O_8$	97.72	645.2482	645.2487	-2.565	293.2123; 235.1705; 275.2013; 95.0137	-H
152	Mexicanolide	$C_{27}H_{32}O_7$	99.65	469.2220	469.2221	0.03	83.0491; 55.0546; 283.1749; 133.0858	+H

Peak Number	Identification Component	Molecular Formula	Retention Time (min)	Theoretical Mass (m/z)	Experimental Mass (m/z)	Delta mmu	Main MS/MS Fragments Detected	Mode
153	α-Linolenic acid	$C_{18}H_{30}O_2$	101.03	277.2162	277.2172	1.083	121.0295; 277.2173	-H
154	dancosterol	$C_{35}H_{60}O_6$	104.64	577.4462	577.4370	-2.536	N/A	+H
155	arundinan	$C_{22}H_{22}O_3$	104.94	335.1641	335.1662	2.119	N/A	+H
156	18 β-Glycyrrhetintic Acid	$C_{30}H_{46}O_4$	105.81	469.3312	469.3324	1.164	425.3426; 469.3329; 426.3468; 451.3211; 116.9284	-H
157	Glabrolide	$C_{30}H_{44}O_4$	107.19	467.3155	467.317	1.434	423.3270; 467.3163; 424.3318; 449.3076; 468.3195	-H
158	Palmitic acid	$C_{16}H_{32}O_2$	108.73	255.2318	255.2329	1.133	255.2330; 74.0248; 116.9286	-H
159	2,4,7-Trimethoxy-9,10-dihydrophenanthrene	$C_{17}H_{18}O_3$	110.50	269.1172	269.2115	0.399	N/A	-H
160	Oleanonic acid	$C_{30}H_{46}O_3$	112.30	453.3363	453.3375	1.228	453.3374; 454.3390; 94.9806	-H

*N/A denotes the meaning of the component cannot find the secondary fragment in ion spectrum.

Table S11. Regression model performance of training data and validation data.

		BPNN		RBFNN	
		Training data	Validation data	Training data	Validation data
Ethanol Extraction Process	Yield of Dry Extraction	0.9996	0.9999	0.9964	0.9987
	Extraction Yield of Schisandrin	0.9966	0.9901	0.9977	0.9986
Water Extraction Process	Yield of Dry Extraction	0.9955	0.9895	0.9947	0.9949
	Content of Total Sugar	0.9984	0.9985	0.9948	0.9694
	Content of Hesperidin	0.9971	0.9997	0.9969	0.9041
	W_{ave}	0.8771	0.8955	0.9764	0.9631
	$\overline{A_{ave}}$	0.9647	0.9839	0.9708	0.9720
	$\overline{Mw_{ave}}$	0.9647	0.9742	0.9636	0.9592

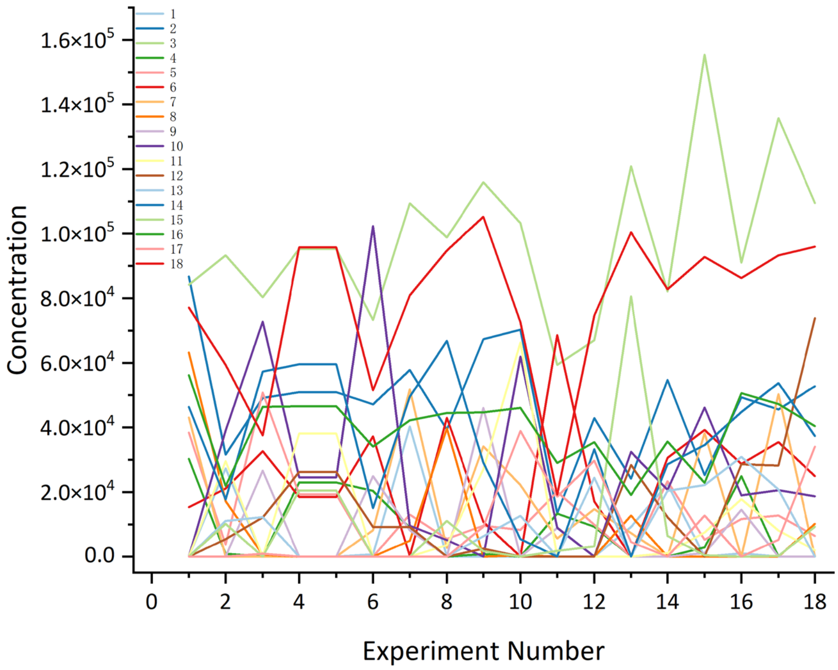

Fig. S5. Concentrations of OED samples calculated by MCR-ALS.

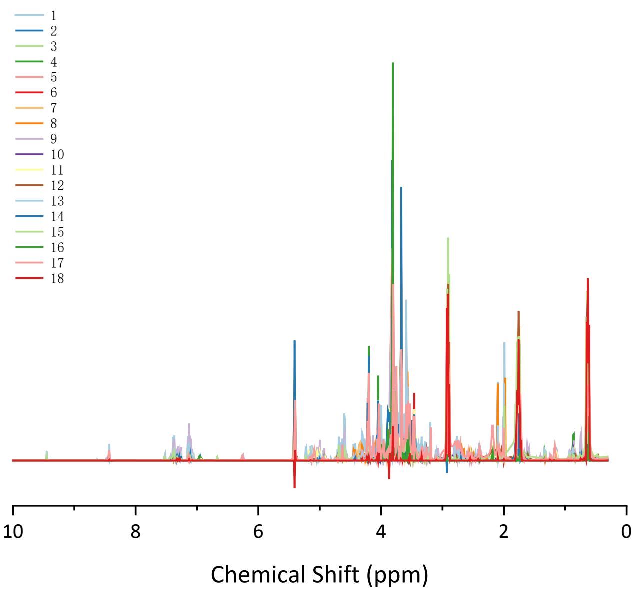

Fig. S6. qNMR spectrum of OED samples deconvoluted by MCR-ALS. by MCR-ALS.

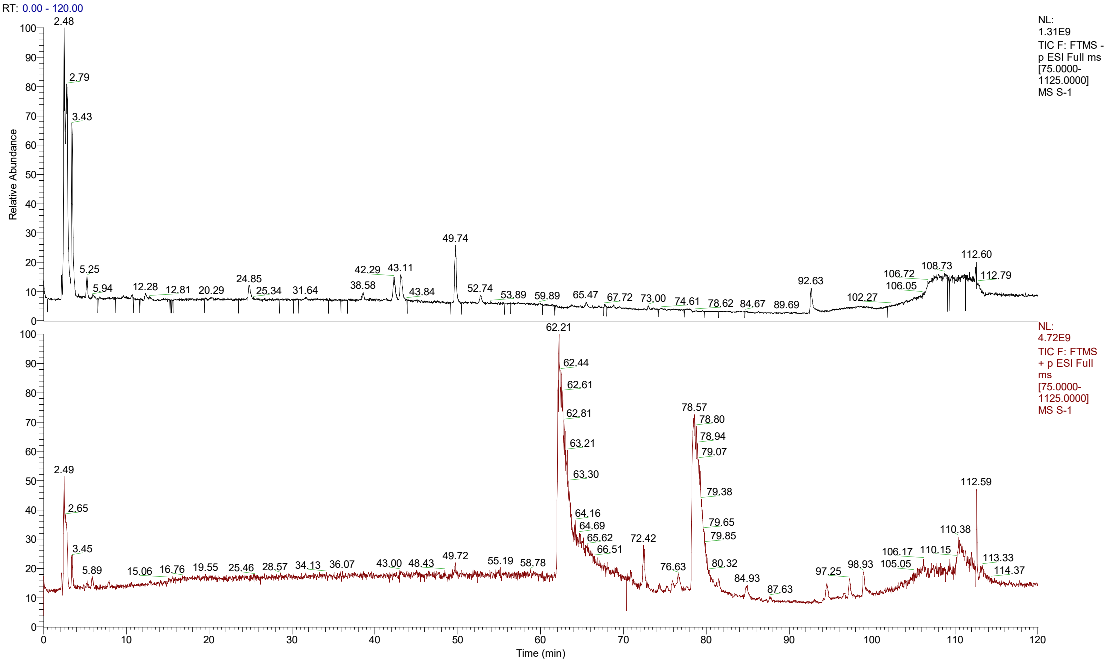

Fig. S7. Total ion chromatogram of ideal extract of BWG on positive and negative mode.

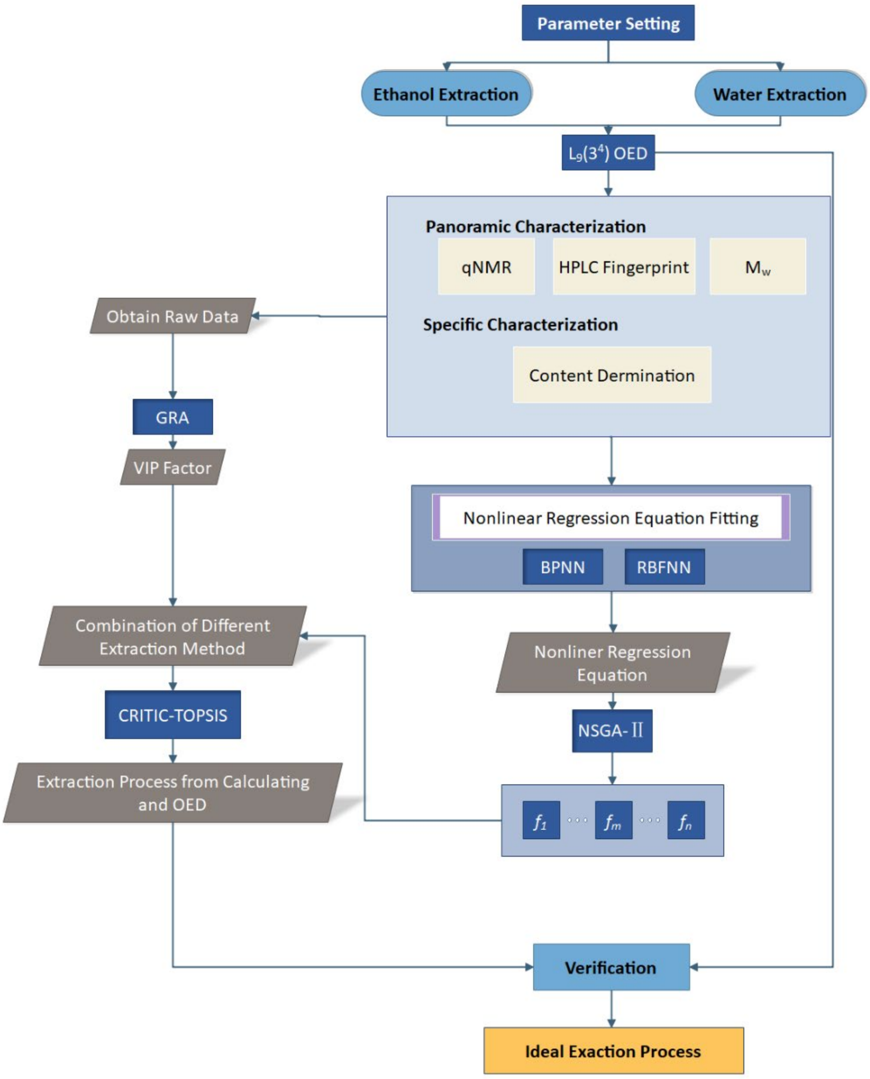

Fig. S8. The workflow of the model to optimize the multiple stage extraction process.

4. References

1 T. Yamazaki, S. Nakamura and T. Saito, *Metrologia*, 2017, **54**, 224–228.

2 S. Westwood, T. Yamazaki, T. Huang, B. Garrido, I. Ün, W. Zhang, G. Martos, N. Stoppacher, T. Saito and R. Wielgosz, *Metrologia*, 2019, **56**, 064001.

3 T.-Q. Peng, X.-L. Yin, H.-W. Gu, W. Sun, B. Ding, X.-C. Hu, L.-A. Ma, S.-D. Wei, Z. Liu and S.-Y. Ye, *Food Chemistry*, 2021, **347**, 128959.

4 D. J. Strydom, *Journal of Chromatography A*, 1994, **678**, 17–23.

5 X. Li, X. Wang, Y. Dong, R. Song, J. Wei, A. Yu, Q. Fan, J. Yao, D. Shan, F. Lv, X. Zhong and G. She, *Industrial Crops and Products*, 2022, **175**, 114288.

6 S. K. Bharti and R. Roy, *TrAC Trends in Analytical Chemistry*, 2012, **35**, 5–26.

7 S. Zhou, G. Huang and G. Chen, *FOOD CHEMISTRY*, , DOI:10.1016/j.foodchem.2021.130089.

8 C. Peng, Y. Zhu, F. Yan, Y. Su, Y. Zhu, Z. Zhang, C. Zuo, H. Wu, Y. Zhang, J. Kan and D. Peng, *Food Chemistry*, 2021, **340**, 127907.

9 L. Zhu, G. Wang, F. Huang, Y. Li, W. Chen and H. Hong, *IEEE Geosci. Remote Sensing Lett.*, 2022, **19**, 1–5.

10 N. Liu, W. Yang, X. Li, P. Zhao, Y. Liu, L. Guo, L. Huang and W. Gao, *FOOD CHEMISTRY*, , DOI:10.1016/j.foodchem.2022.132683.

11 M. Yu, *Carbohydrate Polymers*, 2021, 9.

12 F. Dranca, M. Vargas and M. Oroian, *Food Hydrocolloids*, 2020, **100**, 105383.

13 X. Chen, G. Chen, Z. Wang and J. Kan, *International Journal of Biological Macromolecules*, 2020, **151**, 635–649.

14 X. Li, L. Wang, Y. Wang and Z. Xiong, *PROCESS BIOCHEMISTRY*, 2016, **51**, 1100–1108.