Electronic Supplementary Material (ESI) for RSC Advances.
This journal is © The Royal Society of Chemistry 2019

A proposal protocol based on integrative metabonomics analysis for the rapid discrimination and mechanism explanation of sulfur fumigated Chinese herbal medicines

Dai Shengyun[1,2], Wang Yuqi[1], Wang Fei[1,3], Mei Xiaodan[1], Zhang Jiayu[4,5,*]

[1] *School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 102488, China*

[2] *National Institute of Food and Drug Control, Beijing 100050, China*

[3] *Department of Pharmacy, People Hospital of Peking University, Beijing 100044, China*

[4] *Beijing Research Institute of Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China*

[5] *School of pharmacy, Binzhou Medical University, Yantai 264003, China;*

[*]Correspondence and requests for materials should be addressed to Zhang Jiayu. (email: zhangjiayu0615@163.com)

Supplementary material

Data-acquisition method

HPLC-DAD conditions

The chromatographic separation was achieved on an Agilent 1260 Infinity II system (Agilent, USA) with Agilent TC-C_{18} column (4.6 × 250 mm, 5 μm; Agilent, USA) at 25°C. The mobile phase consisted of 0.4% phosphoric acid (solvent A) and acetonitrile (solvent B), and the linear gradient was as follows: 0-20 min, 3%-15% B; 20-45 min, 15%-24% B; 45-55 min, 24%-40% B; 55-60 min, 40%-3% B. The flow rate was 1.0 mL·min^{-1} and the injection volume was 10 μL. The DAD detection wavelengths were 238 nm, 254 nm, 280 nm and 330 nm. All the samples were randomly coded and subjected into HPLC system, whose stability was validated by running the QC sample every 5 samples during the data-acquisition process.

NIR conditions

An Antaris Nicolet FT-NIR analyzer (Thermo Fisher Scientific Inc., USA) with integrating sphere diffuse reflectance mode was employed to obtain the NIR spectra at a spectral resolution of 8 cm^{-1} with 32 scans. The wavelength range of spectra was from 10,000 cm^{-1} to 4,000 cm^{-1}. To acquire the representative spectrum, each sample was separately scanned three times and the average spectrum was calculated for the final metabonomics analysis. During the measurements, room temperature was maintained at 25°C.

UHPLC-LTQ-Orbitrap MS conditions

All the sample analyses were performed on a DIONEX Ultimate 3000 UHPLC

system (Thermo Scientific, Bremen, Germany) with an ACQUITY UPLC HSS T3 column (100 mm × 2.1 mm, 1.8 μm; Waters Corp., Milford, MA, USA) at 40°C. A linear gradient elution program was conducted for chromatographic separation with 0.5% formic acid (solvent A) and acetonitrile (solvent B) as follows: 0-2 min, 5%-8% B; 2-7 min, 8%-10% B; 7-12 min, 10%-12% B; 12-15 min, 12%-16% B; 15-24 min, 16%-25% B; 24-26 min, 25%-95% B; 26-29 min, 95% B; 29-30 min, 95%-5% B, 30-33 min, 5% B. The flow rate was 0.3 $mL \cdot min^{-1}$ and the injection volume was 2 μL. The QC sample was run every 5 samples to guarantee the system stability.

High resolution MS analysis was performed on an LTQ-Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with electrospray ionization (ESI) source operating in negative ion mode, and the parameters were set as follows: capillary temperature, 350°C; capillary voltage, 25 V; sheath gas flow rate, 30 arb; aux gas flow rate, 5 arb; spray voltage, 3.0 kV; tube lens, 110 V; scan ranges, *m/z* 100-1,000; collision-induced dissociation (CID) collision energy, 35%. To reduce analysis time and simultaneously trigger more target ions, the dynamic exclusion was selected. Other parameters were set as follows: repeat count, 5; repeat duration, 30 s; exclusion duration, 60 s. Meanwhile, to validate the identified sulfur-containing derivatives, an ultra-high-resolution of mass spectrometry (100,000 FWHM @ 400 *m/z*) in full scan mode was employed to screen the sulfur-containing derivatives.

Figure captions:

Figure S1 The plane view of the reference chemical structures.

Figure S2 Typical NIR, HPLC-DAD and total ion chromatogram (TIC) of the representative SF and NSF samples.

Figure S3 The preprocess method for the NIR method for the PCA.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(11) plus first-order derivatives (G), SG(11) plus second-order derivatives (H), standard normal variate transformation (I), and wavelet denosing of spectra (J).

Figure S4 The preprocess method for the NIR method for the PLS-DA.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (MSC, C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(11) plus first-order derivatives (G), SG(11) plus second-order derivatives (H), standard normal variate transformation (I), and wavelet denosing of spectra (J).

Figure S5 The fragmentation pathway of 5-CQA.

Figure S6 Histogram of signal intensity of sulfur derivatives.

Figure S7 The PCA results for the other SiPLS.

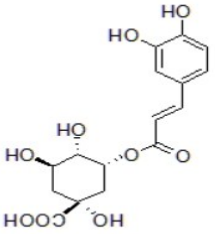
Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus

first-order derivatives (F), SG(9) plus second-order derivatives (H), SG(11) plus first-order derivatives (G), standard normal variate transformation (I), and wavelet denosing of spectra (J).

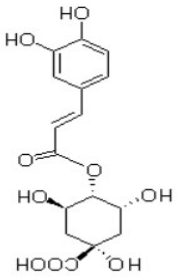
Figure S8 The PLS-DA results for the other SiPLS.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(9) plus second-order derivatives (H), SG(11) plus first-order derivatives (G), standard normal variate transformation (I), and wavelet denosing of spectra (J).

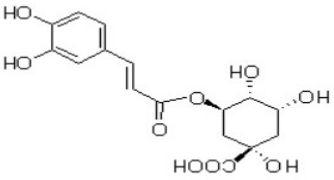
Figure S9 The 2D-COS plot for the SF (A) and NSF (B) samples.



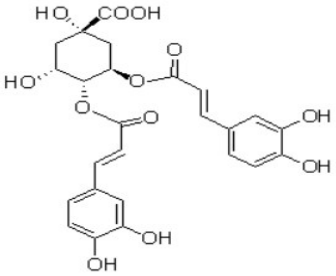
3-CQA



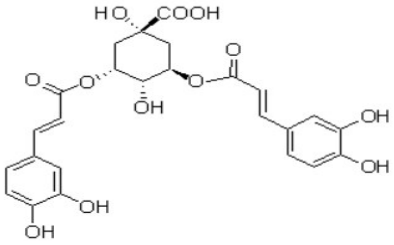
4-CQA



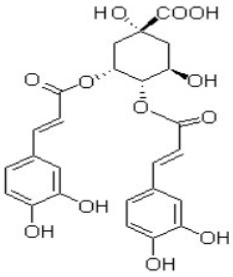
5-CQA



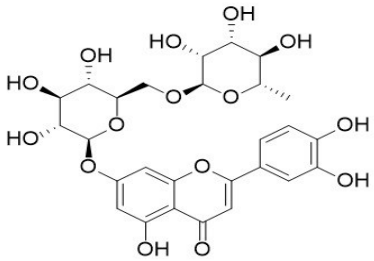
3,4-DiCQA



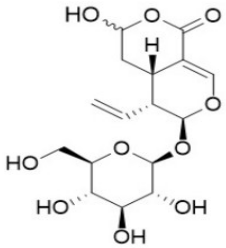
3,5-DiCQA



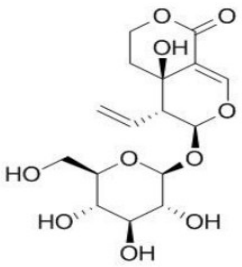
4,5-DiCQA



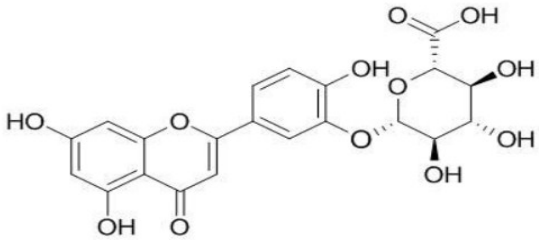
Lonicerin



Secologanic acid

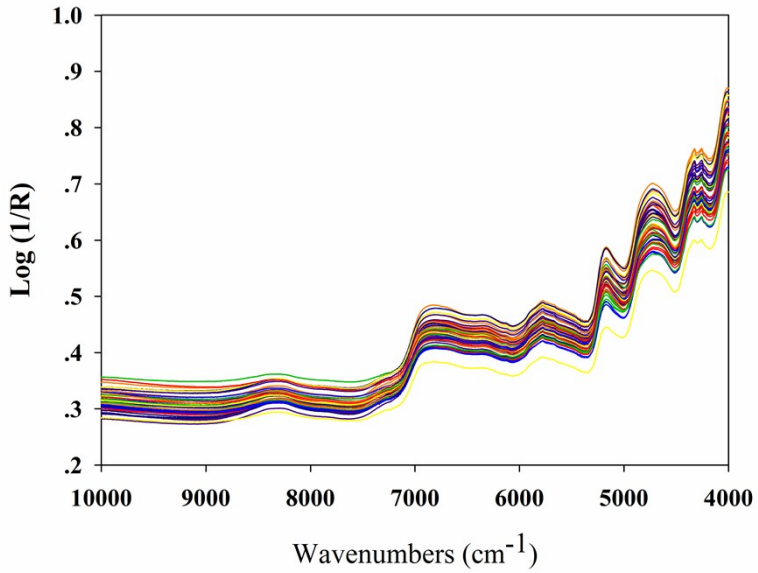


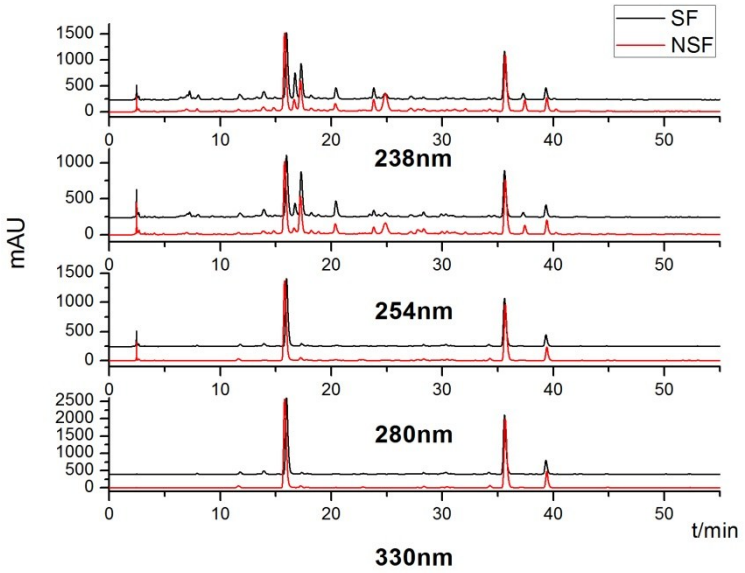
Swertiamarin



Luteolin 7-O-β-glucoside

Figure S1 The plane view of the reference chemical structures.


1.0
.9
.8
.7
.6
.5
.4
.3
.2
Log (1/R)
10000
9000
8000
7000
6000
5000
4000
Wavenumbers (cm-1)


SF
NSF
238nm
254nm
280nm
330nm
mAU
t/min

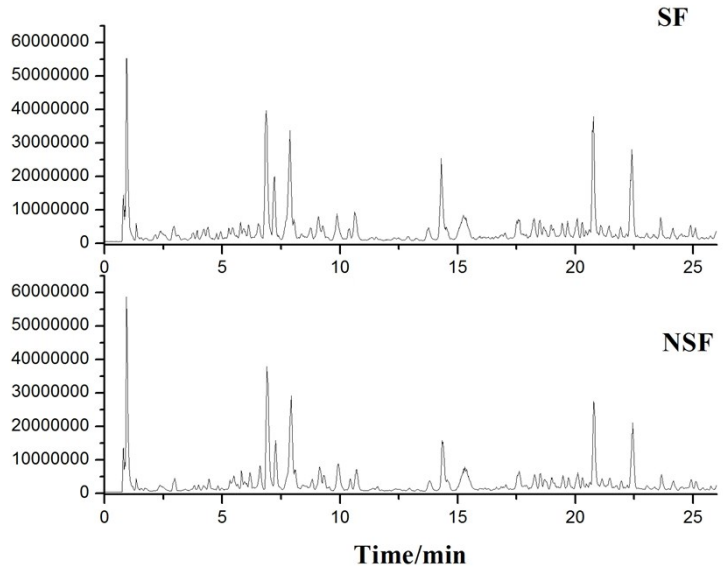
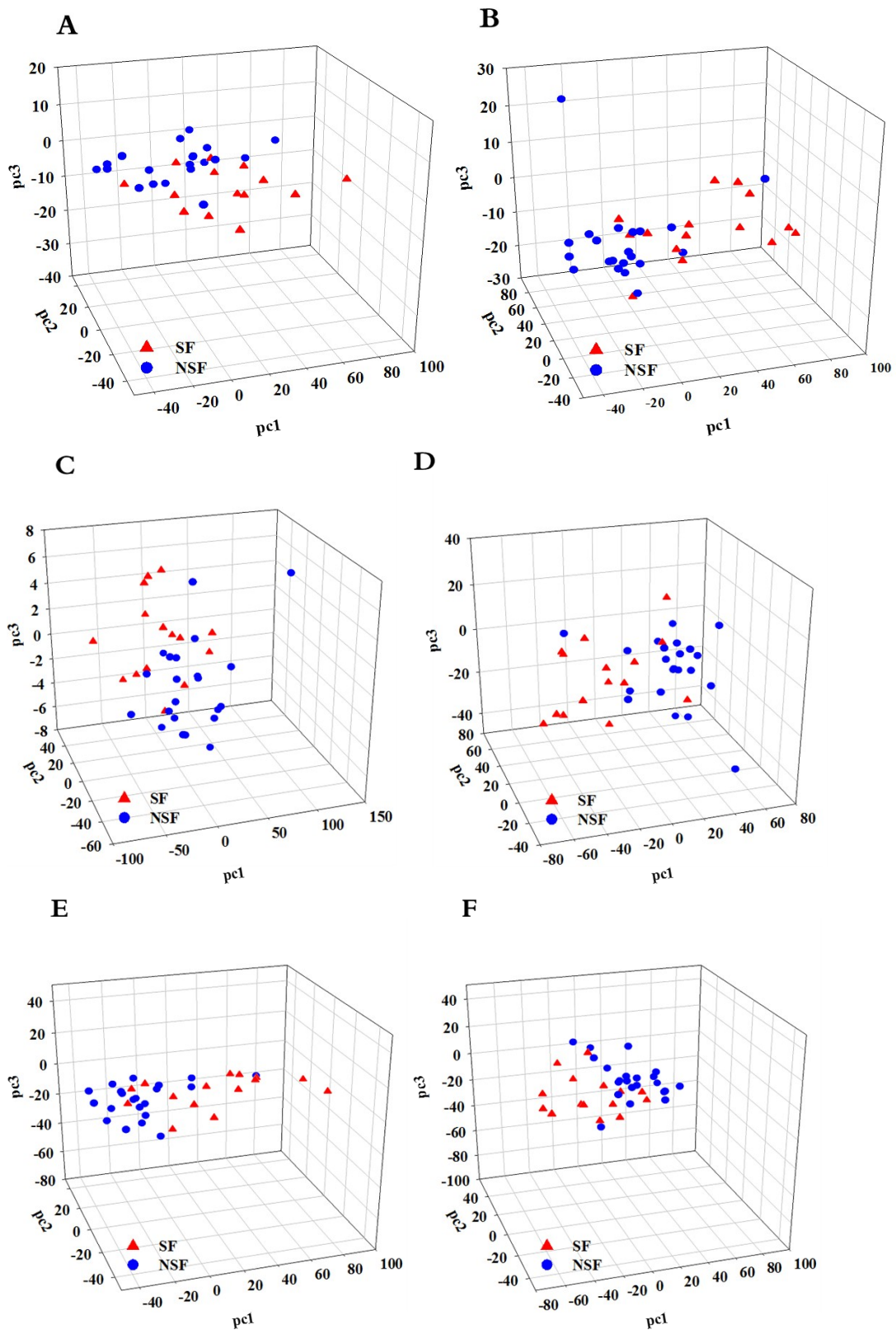


Figure S2 Typical NIR, HPLC-DAD and total ion chromatogram (TIC) of the representative SF and NSF samples.


A
pc3
pc2
pc1
SF
NSF
B
pc3
pc2
pc1
SF
NSF
C
pc3
pc2
pc1
SF
NSF
D
pc3
pc2
pc1
SF
NSF
E
pc3
pc2
pc1
SF
NSF
F
pc3
pc2
pc1
SF
NSF

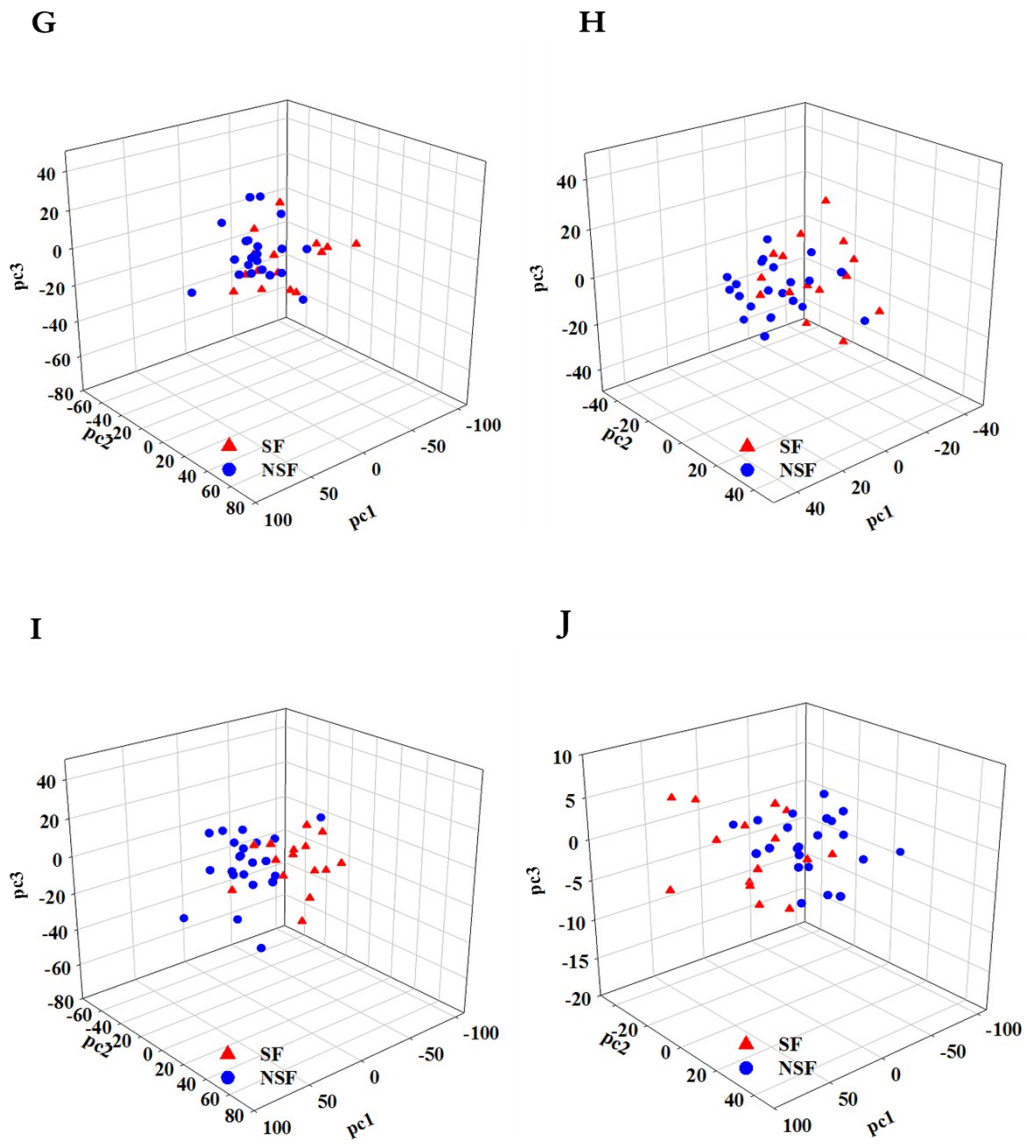
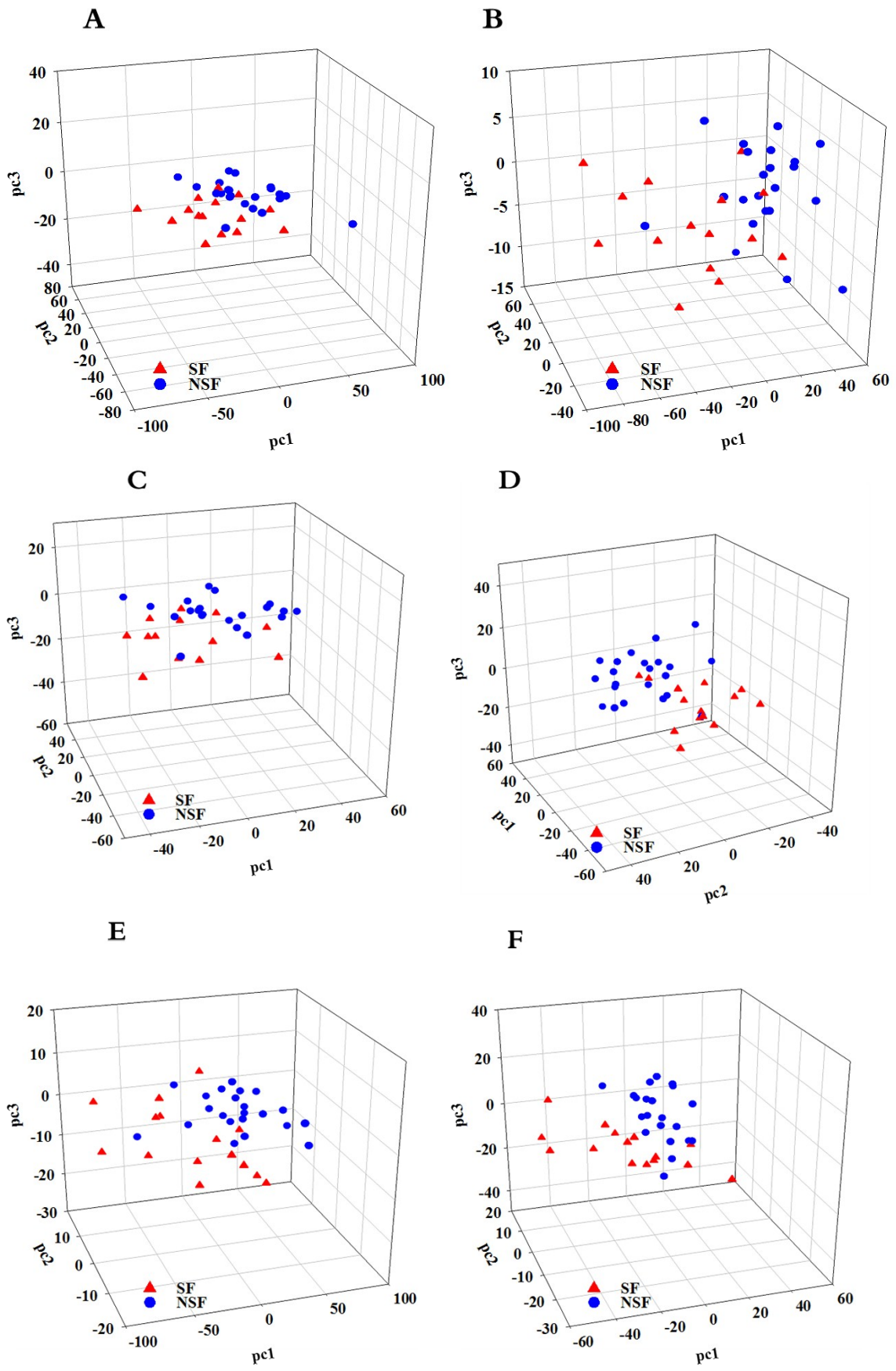


Figure S3 The preprocess method for the NIR method for the PCA.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(11) plus first-order derivatives (G), SG(11) plus second-order derivatives (H), standard normal variate transformation (I), and wavelet denosing of spectra (J).


A
pc3
pc2
pc1
SF
NSF
B
pc3
pc2
pc1
SF
NSF
C
pc3
pc2
pc1
SF
NSF
D
pc3
pc1
pc2
SF
NSF
E
pc3
pc2
pc1
SF
NSF
F
pc3
pc2
pc1
SF
NSF

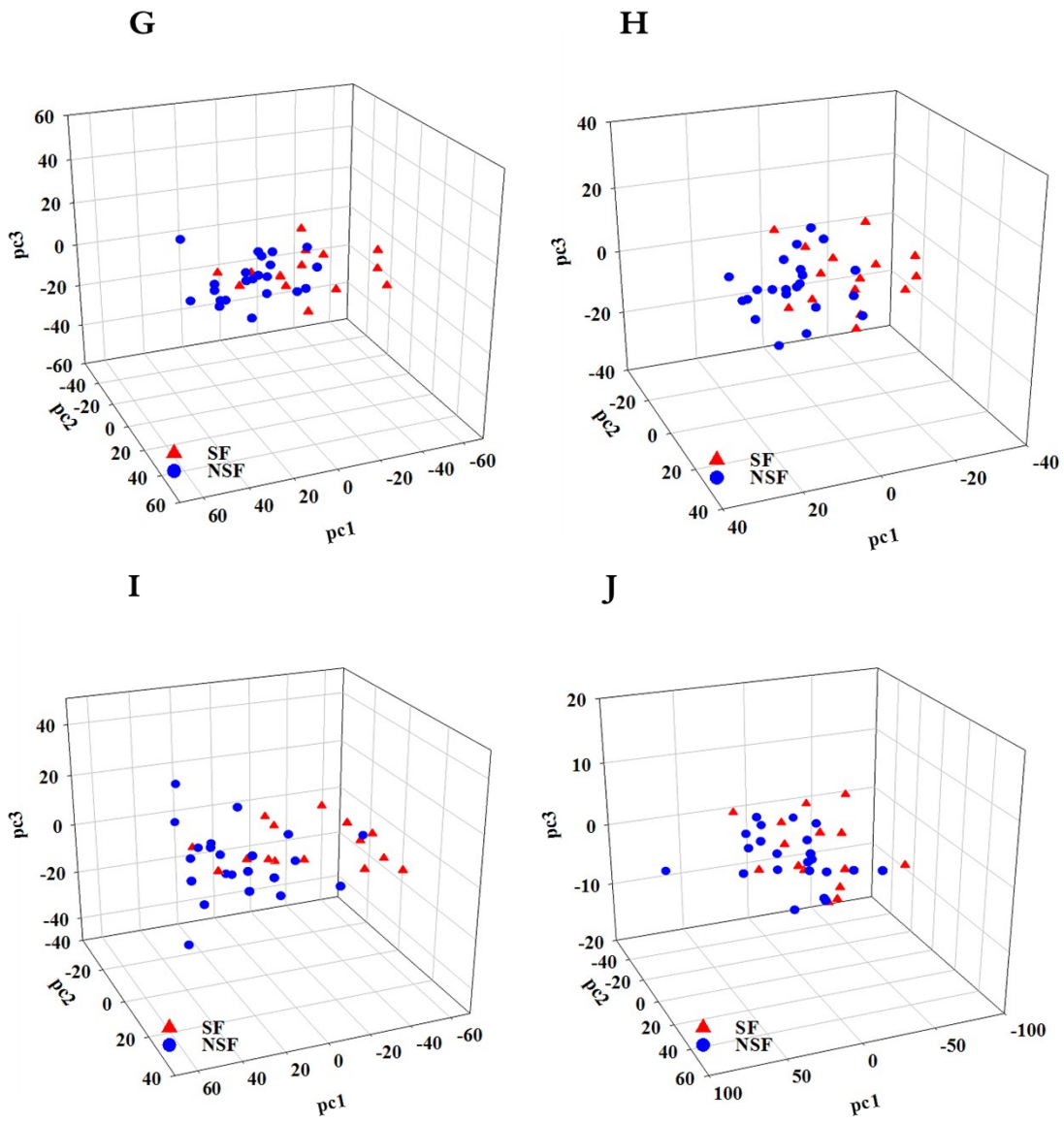


Figure S4 The preprocess method for the NIR method for the PLS-DA.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (MSC, C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(11) plus first-order derivatives (G), SG(11) plus second-order derivatives (H), standard normal variate transformation (I), and wavelet denosing of spectra (J).

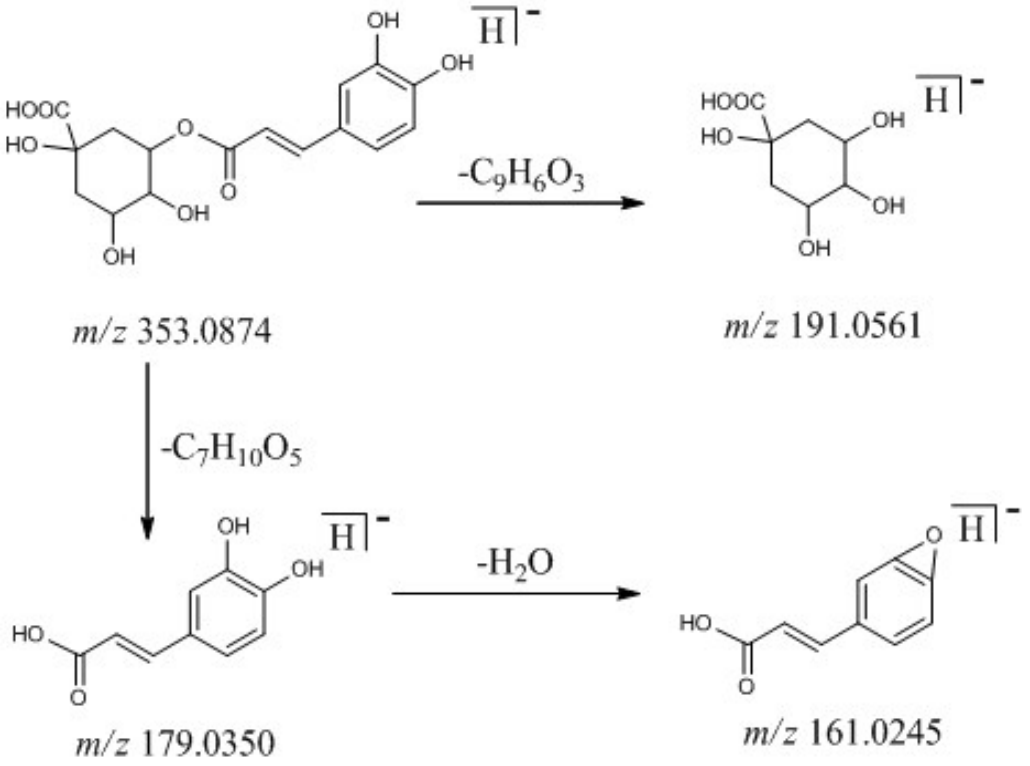


Figure S5 The fragmentation pathway of 5-CQA.

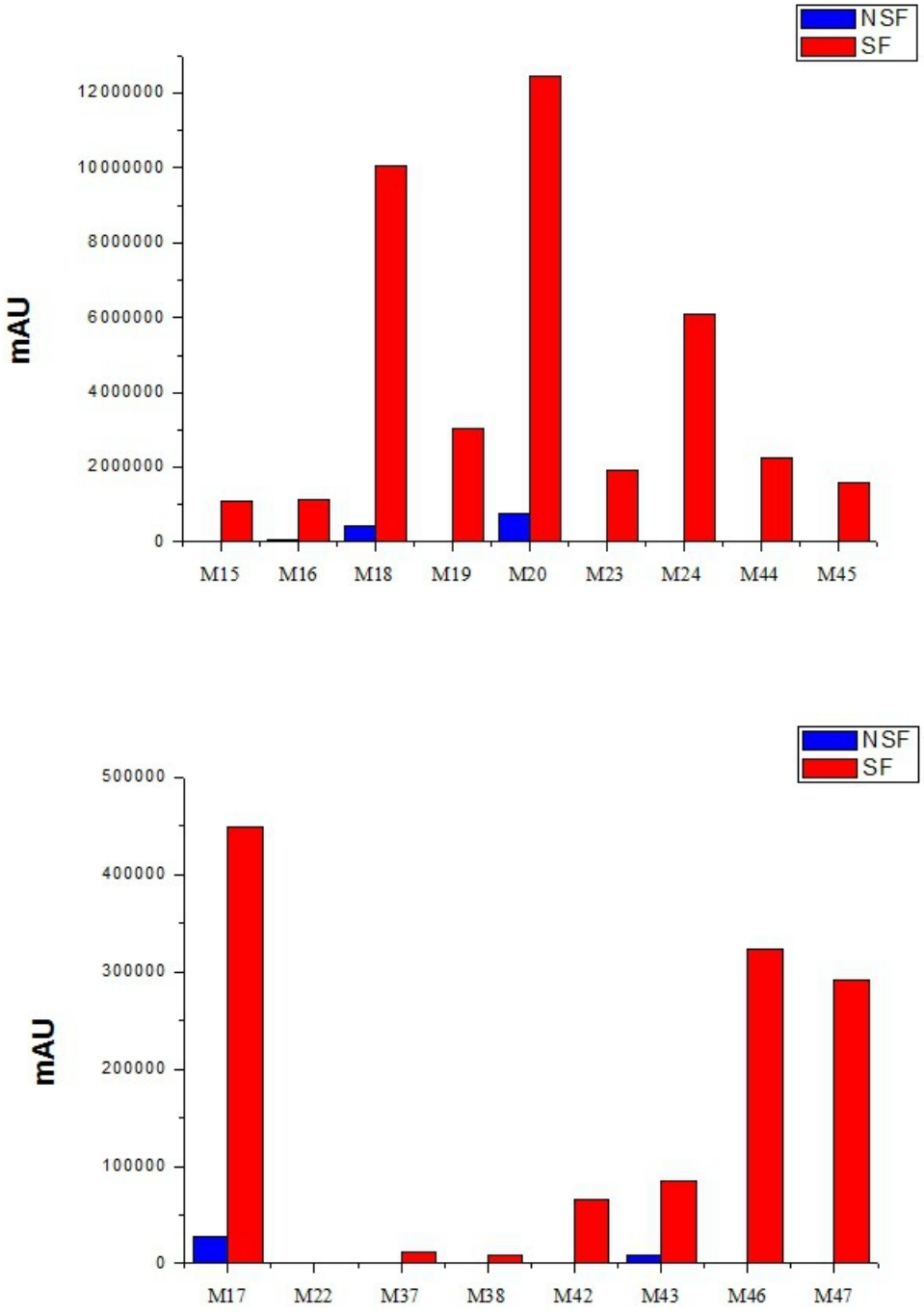
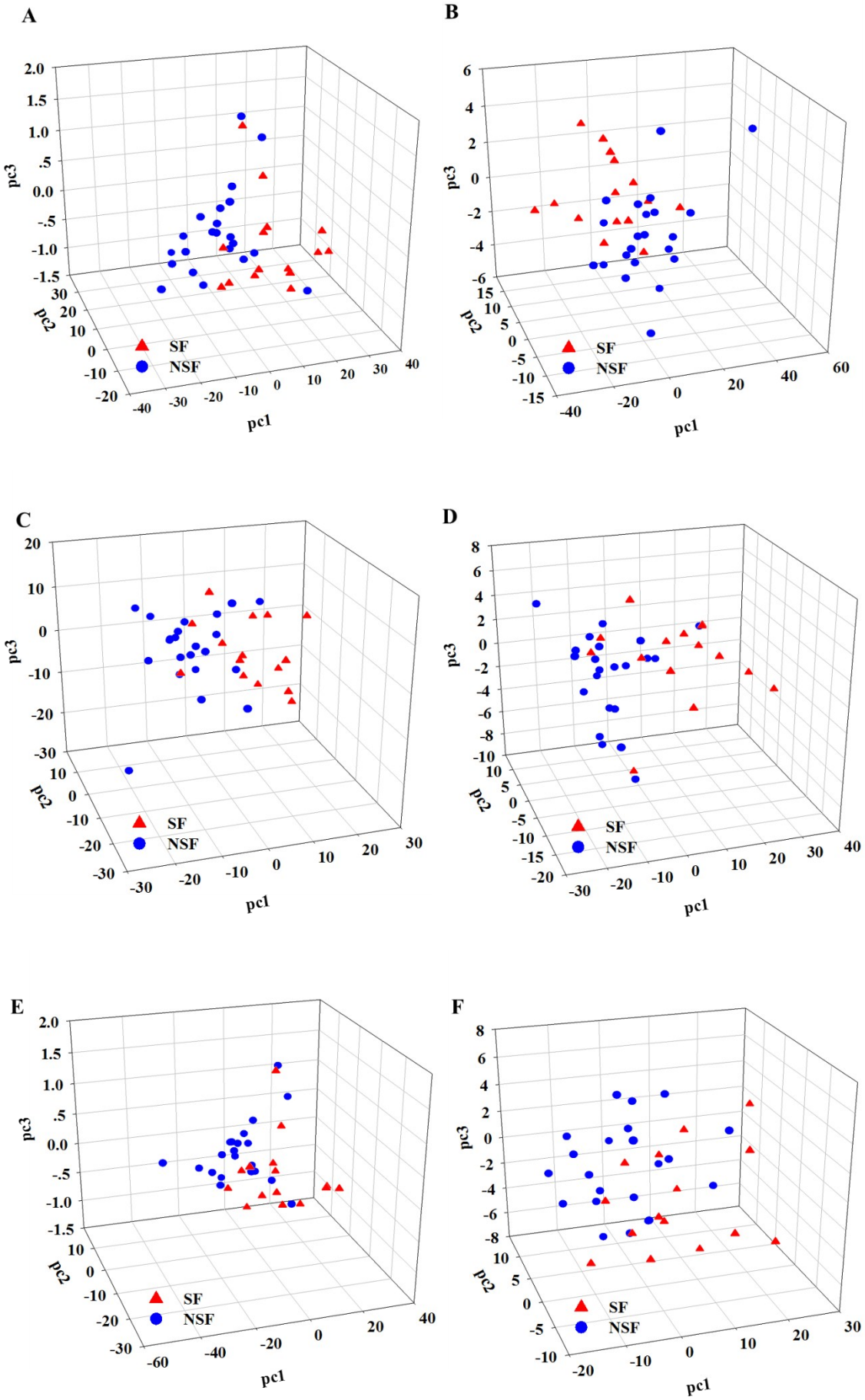


Figure S6 Histogram of signal intensity of sulfur derivatives.


A
pc3
pc2
pc1
SF
NSF
B
pc3
pc2
pc1
SF
NSF
C
pc3
pc2
pc1
SF
NSF
D
pc3
pc2
pc1
SF
NSF
E
pc3
pc2
pc1
SF
NSF
F
pc3
pc2
pc1
SF
NSF

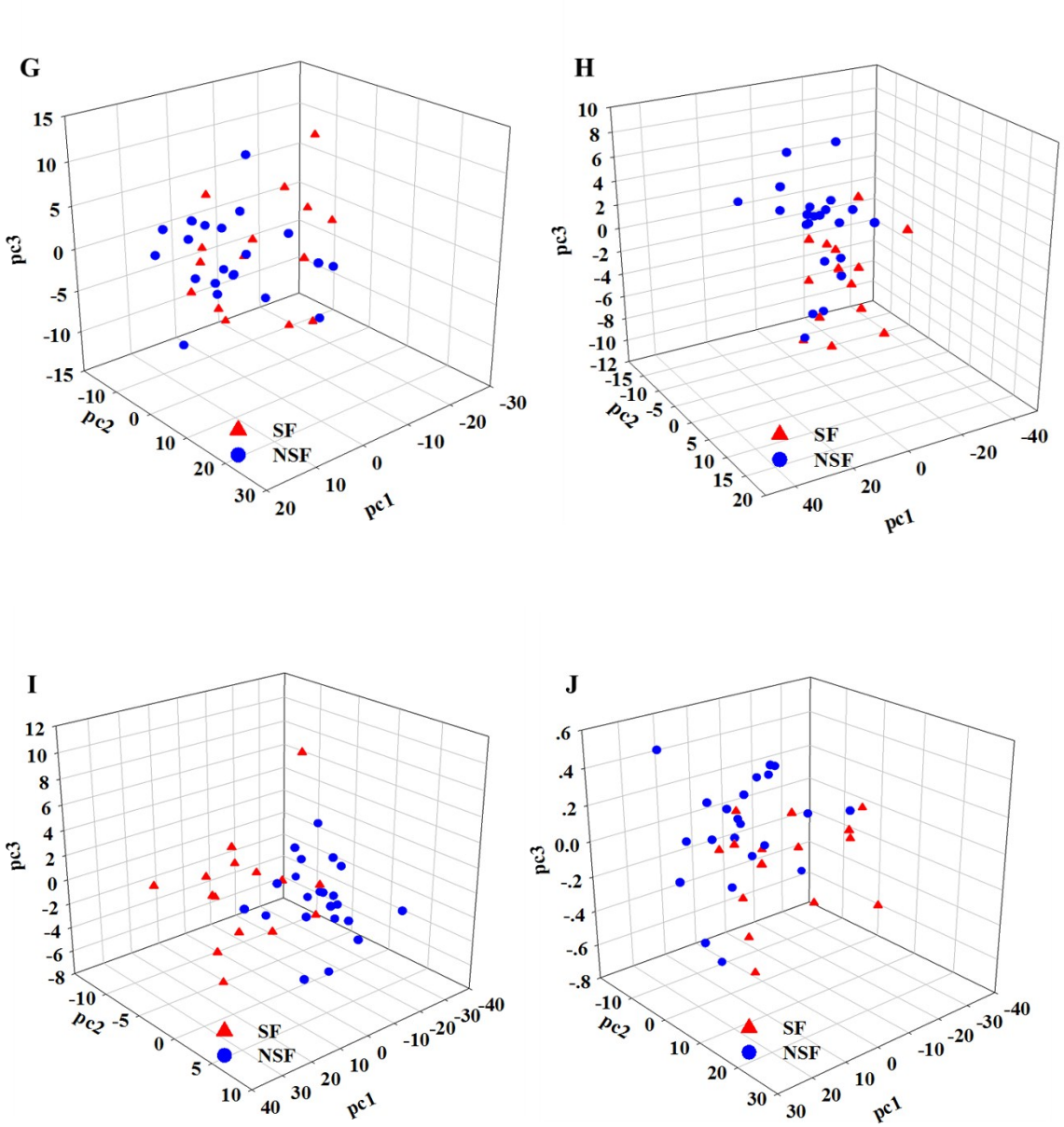
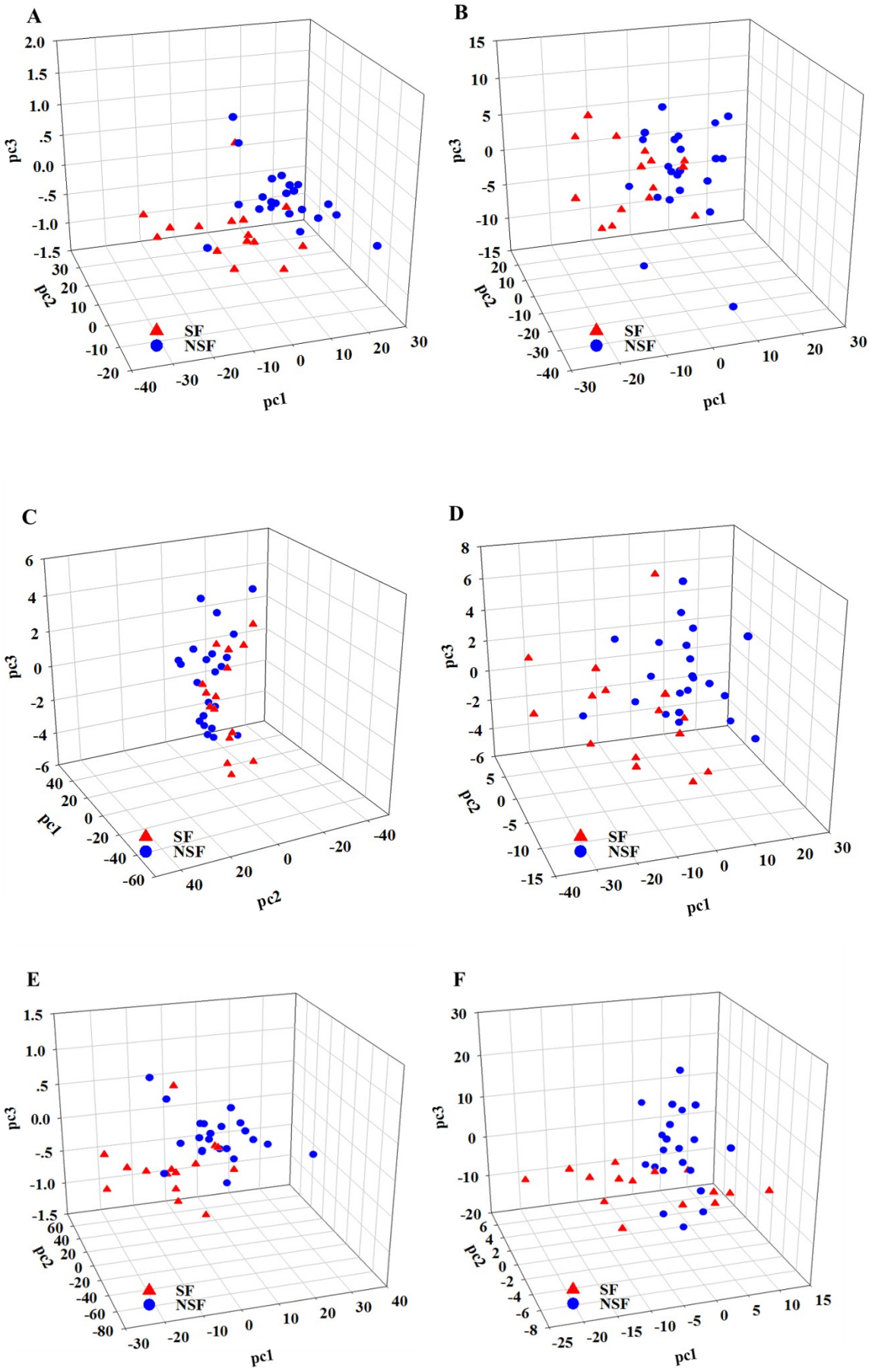


Figure S7 The PCA results for the other SiPLS.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(9) plus second-order derivatives (H), SG(11) plus first-order derivatives (G), standard normal variate transformation (I), and wavelet denosing of spectra (J).


A
pc3
pc2
pc1
SF
NSF
B
pc3
pc2
pc1
SF
NSF
C
pc3
pc1
pc2
SF
NSF
D
pc3
pc2
pc1
SF
NSF
E
pc3
pc2
pc1
SF
NSF
F
pc3
pc2
pc1
SF
NSF

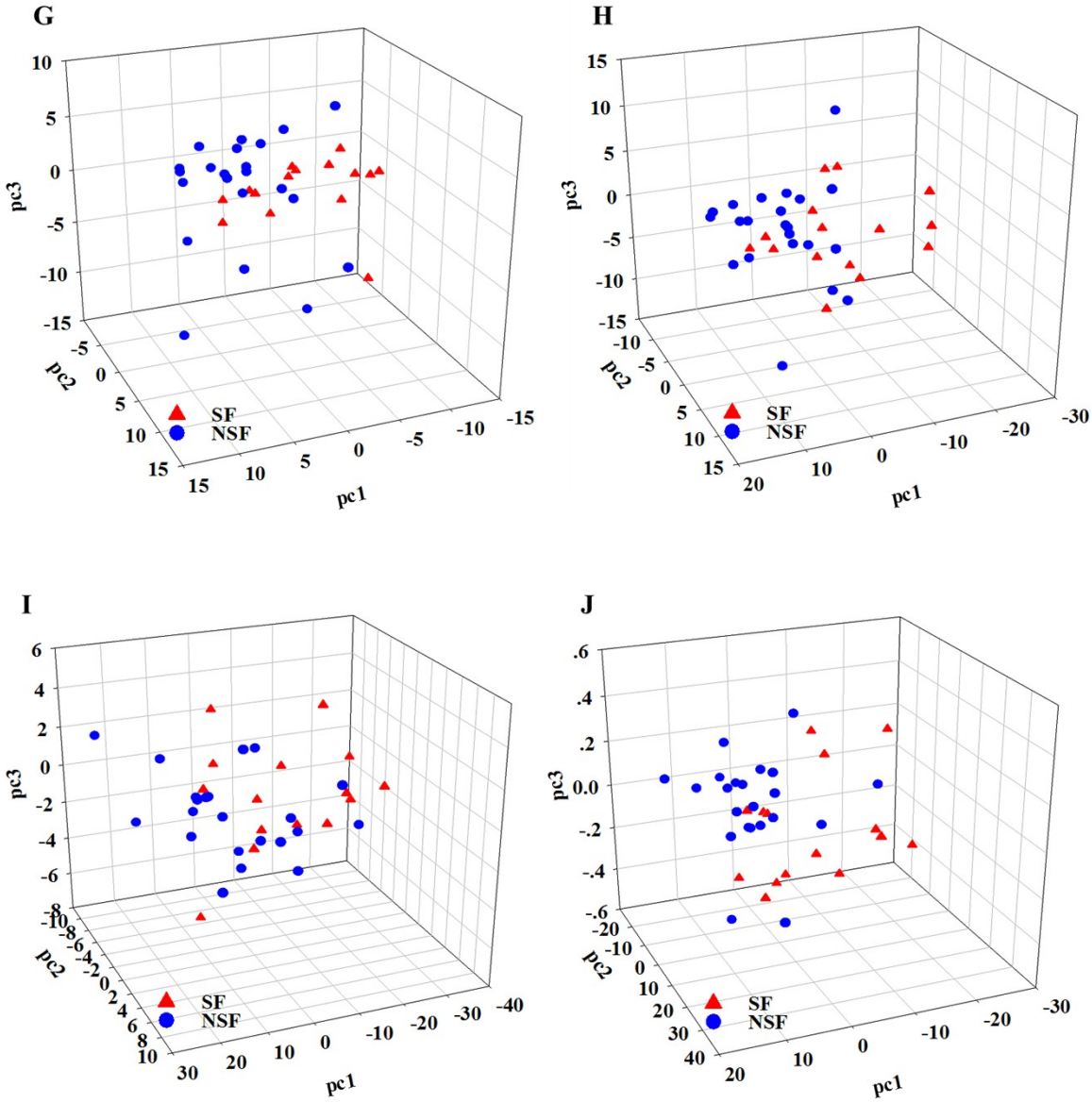


Figure S8 The PLS-DA results for the other SiPLS.

Baseline (A), spectroscopic transformation (B), multiplicative scatter correction (C), normalization (D), original (E), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (F), SG(9) plus second-order derivatives (H), SG(11) plus first-order derivatives (G), standard normal variate transformation (I), and wavelet denosing of spectra (J).

A B

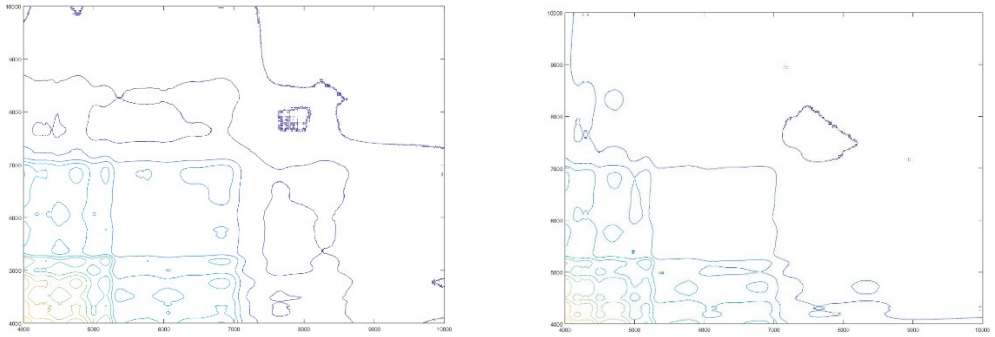


Figure S9 The 2D-COS plot for the SF (A) and NSF (B) samples.

Table captions:

Table S1 The detailed information of raw materials in commercial.

Table S2 The reference standard information.

Table S3 The results of preprocess method of the NIR.

Table S4 The information for the 49 identified markers.

Table S1 The detailed information of raw materials in commercial.

No.	Code No.	Location
1	FLJ-01-1	Mixian city, Henan China
2	FLJ-02-1	Fengqiu city, Henan, China
3	FLJ-03-1	Fengqiu city, Henan, China
4	FLJ-04-1	Fengqiu city, Henan, China
5	FLJ-05-1	Fengqiu city, Henan, China
6	FLJ-06-1	Fengqiu city, Henan, China
7	FLJ-07-1	Fengqiu city, Henan, China
8	FLJ-08-1	Fengqiu city, Henan, China
9	FLJ-09-1	Pingyi city, Shandong, China
10	FLJ-10-1	Pingyi city, Shandong, China
11	FLJ-11-1	Pingyi city, Shandong, China
12	FLJ-12-1	Pingyi city, Shandong, China
13	FLJ-13-1	Pingyi city, Shandong, China
14	FLJ-14-1	Linyi city, Shandong, China
15	FLJ-15-1	Linyi city, Shandong, China
16	FLJ-16-1	Beijing city, Beijing, China
17	FLJ-17-1	Beijing city, Beijing, China
18	FLJ-18-1	Guangzhou city, Guangdong, China
19	FLJ-19-1	Guangzhou city, Guangdong, China

20	FLJ-20-1	Guangzhou city, Guangdong, China
21	FLJ-21-1	Guangzhou city, Guangdong, China
22	FLJ-22-1	Nanjing city, Jiangsu, China
23	FLJ-01-2$^{\Delta}$	Mixian city, Henan, China
24	FLJ-02-2$^{\Delta}$	Fengqiu city, Henan, China
25	FLJ-03-2$^{\Delta}$	Fengqiu city, Henan, China
26	FLJ-04-2$^{\Delta}$	Fengqiu city, Henan, China
27	FLJ-05-2$^{\Delta}$	Fengqiu city, Henan, China
28	FLJ-06-2$^{\Delta}$	Fengqiu city, Henan, China
29	FLJ-07-2$^{\Delta}$	Fengqiu city, Henan, China
30	FLJ-08-2$^{\Delta}$	Fengqiu city, Henan, China
31	FLJ-09-2$^{\Delta}$	Pingyi city, Shandong, China
32	FLJ-10-2$^{\Delta}$	Pingyi city, Shandong, China
33	FLJ-11-2$^{\Delta}$	Pingyi city, Shandong, China
34	FLJ-12-2$^{\Delta}$	Pingyi city, Shandong, China
35	FLJ-13-2$^{\Delta}$	Pingyi city, Shandong, China
36	FLJ-14-2$^{\Delta}$	Linyi city, Shandong, China
37	FLJ-15-2$^{\Delta}$	Linyi city, Shandong, China

$^{\Delta}$: Sulfur fumigated.

Table S2 The reference standard information.

Analytes	CAS	Source	Content (%)
3-Caffeoylquinic acid	906-33-2	Chengdu Bio-purify Phytochemicals Ltd	>98.00
4-Caffeoylquinic acid	905-99-7	Chengdu Bio-purify Phytochemicals Ltd	>98.00
5-Caffeoylquinic acid	327-97-9	Chengdu Bio-purify Phytochemicals Ltd	>98.00
3,4-Dicaffeoylquinic acid	14534-61-3	Chengdu Bio-purify Phytochemicals Ltd	>98.00
3,5-Dicaffeoylquinic acid	2450-53-5	Chengdu Bio-purify Phytochemicals Ltd	>98.00
4,5-Dicaffeoylquinic acid	32451-88-0	Chengdu Bio-purify Phytochemicals Ltd	>98.00
Lonicerin	25694-72-8	Chengdu Bio-purify Phytochemicals Ltd	>98.00
Secologanic acid	60077-40-5	Chengdu Bio-purify Phytochemicals Ltd	>98.00
Swertiamarin	17388-39-5	Chengdu Bio-purify Phytochemicals Ltd	>98.00
Luteolin 7-O-β-glucoside	53527-42-7	Chengdu Bio-purify Phytochemicals Ltd	>98.00

Table S3 The results of preprocess method of the NIR

Preprocess	PCA			PLS-DA			
methods	LVs	R^2x	Q^2	lv	R^2x	R^2y	Q^2
Baseline	5	0.995	0.983	3	0.952	0.475	0.275
spectroscopic transformation	4	0.997	0.996	3	0.991	0.491	0.260
msc	9	0.997	0.992	3	0.846	0.545	0.251
normalization	6	0.998	0.996	3	0.952	0.507	0.293
original	4	0.997	0.996	3	0.991	0.495	0.267
SG91st	7	0.778	0.609	3	0.551	0.797	0.237
SG92nd	5	0.39	0.109	3	0.228	0.972	0.552
SG111st	7	0.817	0.673	3	0.569	0.764	0.221
SG112nd	5	0.425	0.154	4	0.307	0.990	0.639
snv	8	0.996	0.989	3	0.848	0.547	0.250
wds	4	0.998	0.997	3	0.992	0.474	0.237

Multiplicative scatter correction (MSC), Savitzky-Golay smoothing with 9 points (SG(9)) plus first-order derivatives (SG 91st), SG(11) plus first-order derivatives (SG 92nd), SG(11) plus first-order derivatives (SG 111st), SG(11) plus second-order derivatives (SG 112nd), standard normal variate transformation (SNV), and wavelet denosing of spectra (WDS).

LVs: the number of latent variables.

Table S4 The information for the 49 identified markers.

No	t_R	Experimental Mass	Relative intensity	Formular $[M-H]^-$	MS/MS fragment ions	Identification
1	2.14	373.1122	4.36×10^6	$C_{16}H_{21}O_{10}$	MS^2[373]:193,149,167 ,179,119	Swertiamarin
2#	5.30	373.1118	1.23×10^6	$C_{16}H_{21}O_{10}$	MS^2[373]:211,167,149,193,179	Secologanic acid
3	7.88	373.1118	9.69×10^6	$C_{16}H_{21}O_{10}$	MS^2[373]:193,149,167,179	Swertiamarin isomer
4	3.15	437.0720	3.15×10^6	$C_{16}H_{21}O_{12}S$	MS^2[437]:193,149,373,355	Secologanic acid+SO_2
5	1.93	455.0836	3.35×10^6	$C_{16}H_{23}O_{13}S$	MS^2[455]:373,411,437,193,211	Secologanic acid +H_2SO_3
6	2.15	455.0822	5.08×10^6	$C_{16}H_{23}O_{13}S$	MS^2[455]:373,437,411,193,211	Secologanic acid +H_2SO_3
7	1.81	391.1231	1.64×10^5	$C_{16}H_{23}O_{11}$	MS^2[391]:229,211,193,185,167,149	Secologanic acid +H^+/H_2O
8	2.45	391.1255	8.88×10^5	$C_{16}H_{23}O_{11}$	MS^2[391]:211,229,193,167,149,185	Secologanic acid +H^+/H_2O
9#	4.47	353.0869	2.34×10^6	$C_{16}H_{17}O_9$	MS2[353]:191,179,135	3-CQA
10#	6.91	353.0858	1.81×10^7	$C_{16}H_{17}O_9$	MS^2[353]:191,179,161	5-CQA
11#	7.73	353.0856	1.24×10^6	$C_{16}H_{17}O_9$	MS^2[353]:173,179,191,135	4-CQA
12	4.23	375.1292	1.242×10^6	$C_{16}H_{23}O_{10}$	MS^2[375]:213,169,151	Loganin acid isomer
13#	4.84	375.1280	6.51×10^5	$C_{16}H_{23}O_{10}$	MS^2[375]:213,169,151,195	Loganin acid isomer
14	5.84	375.1273	1.67×10^6	$C_{16}H_{23}O_{10}$	MS^2[375]:213,169,151	Loganin acid isomer
15	6.63	375.1292	$23.\times10^6$	$C_{16}H_{23}O_{10}$	MS^2[375]:195,151,	Loganin acid isomer

16#	14.33	403.1223	7.88×10^6	$C_{17}H_{23}O_{11}$	MS^2[403]: 371,223,179,121,91	Secologanin
17	1.63	433.0428	5.21×10^6	$C_{16}H_{17}O_{12}S$	MS^2[433]:241,415,353,161,191,287	CQA+SO_3
18	2.53	433.0427	1.60×10^6	$C_{16}H_{17}O_{12}S$	MS^2[433]:415,387,353,241,353	CQA+SO_3
19	2.66	433.0433	1.90×10^6	$C_{16}H_{17}O_{12}S$	MS^2[433]:241,415,387,259,353	CQA+SO_3
20	4.62	433.0423	3.36×10^6	$C_{16}H_{17}O_{12}S$	MS^2[433]:415.387,259	CQA+SO_3
21	5.01	433.0419	4.98×10^6	$C_{16}H_{17}O_{12}S$	MS^2[433]:415,241,161,259,387	CQA+SO_3
22	1.12	435.0591	2.40×10^5	$C_{16}H_{17}O_{12}S$	MS^2[435]:353,191,179	CQA+H_2SO_3
23#	19.06	447.0918	2.48×10^6	$C_{21}H_{19}O_{11}$	MS^2[447]:285	Luteolin-7-*O*-glucoside
24	21.06	447.0916	5.90×10^5	$C_{21}H_{19}O_{11}$	MS^2[447]:285	Luteolin-7-*O*-glucoside isomer
25	18.22	463.0861	1.38×10^6	$C_{21}H_{19}O_{12}$	MS^2[463]:301,271,445	Hyperoside isomer
26#	18.73	463.0854	2.01×10^6	$C_{21}H_{19}O_{12}$	MS^2[463]:301,445,271	Hyperoside
27	23.05	499.1231	9.51×10^4	$C_{25}H_{23}O_{11}$	MS^2[499]:337,173,335,353	4-*P*Co-1-CQA
28	23.49	499.1233	1.11×10^5	$C_{25}H_{23}O_{11}$	MS^2[499]:353,337,191,335,179	5-*P*Co-3-CQA
29	25.21	499.1230	1.73×10^6	$C_{25}H_{23}O_{11}$	MS^2 [499]:353,337,179,191	3-*P*Co-4-CQA
30#	20.36	515.1155	3.46×10^6	$C_{25}H_{23}O_{11}$	MS^2 [515]:353,335,173,179	3,4-DiCQA
31#	20.85	515.1155	4.38×10^6	$C_{25}H_{23}O_{11}$	MS^2 [515]:353,191,179,335	3,5-DiCQA
32#	22.44	515.1163	7.30×10^6	$C_{25}H_{23}O_{11}$	MS^2 [515]:353,191,179,335,353	4,5-DiCQA

33	16.70	595.0750	1.56×10^5	$C_{25}H_{23}O_{15}S$	MS^2 [595]:549,577,415,241,259	DiCQA+SO_3
34	16.98	595.0748	3.37×10^5	$C_{25}H_{23}O_{15}S$	MS^2 [595]:549,577,415,301,397	DiCQA+SO_3
35	17.61	595.0737	2.62×10^6	$C_{25}H_{23}O_{15}S$	MS^2[595]:577,549,415,433,241,259	DiCQA+SO_3
36	17.89	595.0745	5.84×10^5	$C_{25}H_{23}O_{15}S$	MS^2[595]:577,549,415,433,241,259	DiCQA+SO_3
37	19.38	595.0745	6.70×10^5	$C_{25}H_{23}O_{15}S$	MS^2 [595]:577,549,415,433,259	DiCQA+SO_3
38	21.25	595.0737	6.75×10^5	$C_{25}H_{23}O_{15}S$	MS^2[595]:577,415,549,433,259,241	DiCQA+SO_3
39	23.82	529.1343	5.65×10^4	$C_{26}H_{25}O_{12}$	MS^2[529]:367,179,335,353,193	3-C-4-FQA
40	24.60	529.1340	7.14×10^4	$C_{26}H_{25}O_{12}$	MS^2[529]:353,367,191,179	5-C-3-FQA
41	25.86	529.1335	1.27×10^5	$C_{26}H_{25}O_{12}$	MS^2[529]:353,367,173,335	*Cis*-5-C-3-FQA
42#	18.30	609.14038	2.69×10^6	$C_{27}H_{29}O_{16}$	MS^2[609]:301,300,271,255,179,591	Rutin
43	18.80	593.1488	6.08×10^5	$C_{27}H_{29}O_{15}$	MS^2[593]:285,447	Lonicerin isomer
44	19.71	593.1483	1.51×10^6	$C_{27}H_{29}O_{15}$	MS^2[593]:285,447	Lonicerin isomer
45#	20.50	593.1486	9.56×10^5	$C_{27}H_{29}O_{15}$	MS^2[593]:285	Lonicerin
46#	22.70	607.1653	2.84×10^5	$C_{28}H_{31}O_{15}$	MS^2[607]:299	Chrysoeriol-7-*O*-*β*-D-neohesperidoside
47	8.73	543.0431	2.67×10^5	$C_{21}H_{19}O_{15}S$	MS^2[543]:463,381,525,301	Hyperoside+SO_3
48	12.76	543.0432	8.61×10^4	$C_{21}H_{19}O_{15}S$	MS^2[543]:381,301,381,463	Hyperoside +SO_3

49	17.34	527.0494	1.08×10^5	$C_{21}H_{19}O_{14}S$	MS^2[527]:447, 285,481	Luteolin-7-*O*-glucoside +SO_3