

A Multivariate Statistical Approach using UPLC/TOF-MS for Traditional Herbal Medicine Analysis to Compare Different Sample Classes

Kate Yu, Jose M. Castro-Perez, John P. Shockcor

Waters Corporation

Abstract

This application note demonstrates how to apply MarkerLynx Software for sample profiling using a multivariate statistical approach. As a result, sample comparison can be completed in hours with complete profile information on hand. This significantly reduces analysis time and manpower required for THM sample profiling.

MarkerLynx is now replaced with Progenesis QI.

Introduction

Sample profiling is important for Traditional Herbal Medicine (THM) or Traditional Chinese Medicine (TCM) studies simply because there is very little reproducibility from sample to sample. The contents of plant extracts may vary significantly depending on the plant location, harvest time, as well as the extraction method. One

cannot assume identical contents for two samples even if they were extracted from the same plant, or from two plants having the same name.

In addition, there is also a strong need to compare THM samples for quality control. THM sample profiling is also critical for the study of the THM's physiological working mechanisms.

We have developed a simple and fast generic analytical workflow for THM sample analysis (Figure 1). This workflow takes advantage of Waters UPLC technology for high resolution, high sensitivity, and high-speed separations, as well as the SYMAPT HDMS for its TOF-exact mass measurement capability. This workflow can be adopted for either compound identification or for sample profiling.

Compound identification for THM has been discussed elsewhere.¹ This application note demonstrates how to apply this workflow for sample profiling using a multivariate statistical approach. As a result, sample comparison can be completed in hours with complete profile information on hand. This significantly reduces analysis time and manpower required for THM sample profiling.

Experimental

Two samples of a Chinese Ginseng extract drink were used for this work.

- Sample 1 was the Extra Strong Ginseng Extractum (Made in China, distributed by JV Trading Ltd., New York, NY)
- Sample 2 was the Imperial Wild Panax Ginseng Herb Tea (Made in China, distributed by Overseas Factor Corporation, San Francisco, CA)

Each sample was filtered prior to injection.

LC Conditions

LC system: ACQUITY UPLC System

Column: ACQUITY UPLC HSS T3 Column 2.1 x 100 mm, 1.7

μm, 65 °C

Flow rate: 600 μL/min

Mobile phase A: Water + 0.1% Formic Acid

Mobile phase B: MeOH

Gradient:

Time	Composition	Curve
0 min	95% A	-
10 min	30% A	Curve 6
17 min	0% A	Curve 6
20 min	95% A	Curve 1

MS Conditions

MS system: SYNAPT HDMS System

Ionization mode: Electrospray

Capillary voltage: 3000 V

Cone voltage: 35 V

Desolvation temp.: 450 °C

Desolvation gas:	800 L/Hr
Source temp.:	120 °C
Acquisition range:	50 to 1500 <i>m/z</i>
Collision gas:	Argon

Data Process

Compound screening and profiling:	MarkerLynx Application Manager
Multivariate statistical analysis:	SIMPCA-P

Results and Discussion

To ensure the statistical validity and significance of the results, each sample needed to be injected with no less than three replicates. To obtain the complete profile from a sample, it is necessary to run the LC-MS analysis in both positive and negative modes. For this work, each sample was injected six times: three in ESI+ mode and three in ESI- mode. For demonstration purposes, only the results from ESI- are discussed.

Figure 2 shows the comparison of the two base peak ion (BPI) chromatograms obtained from the two Ginseng extract drinks. It appears as though the Extra Strong Ginseng contains a larger number of components at higher concentrations compared with the Wild Panax Ginseng. Further chemical profiling of the two samples requires the use of multivariate statistical tools.

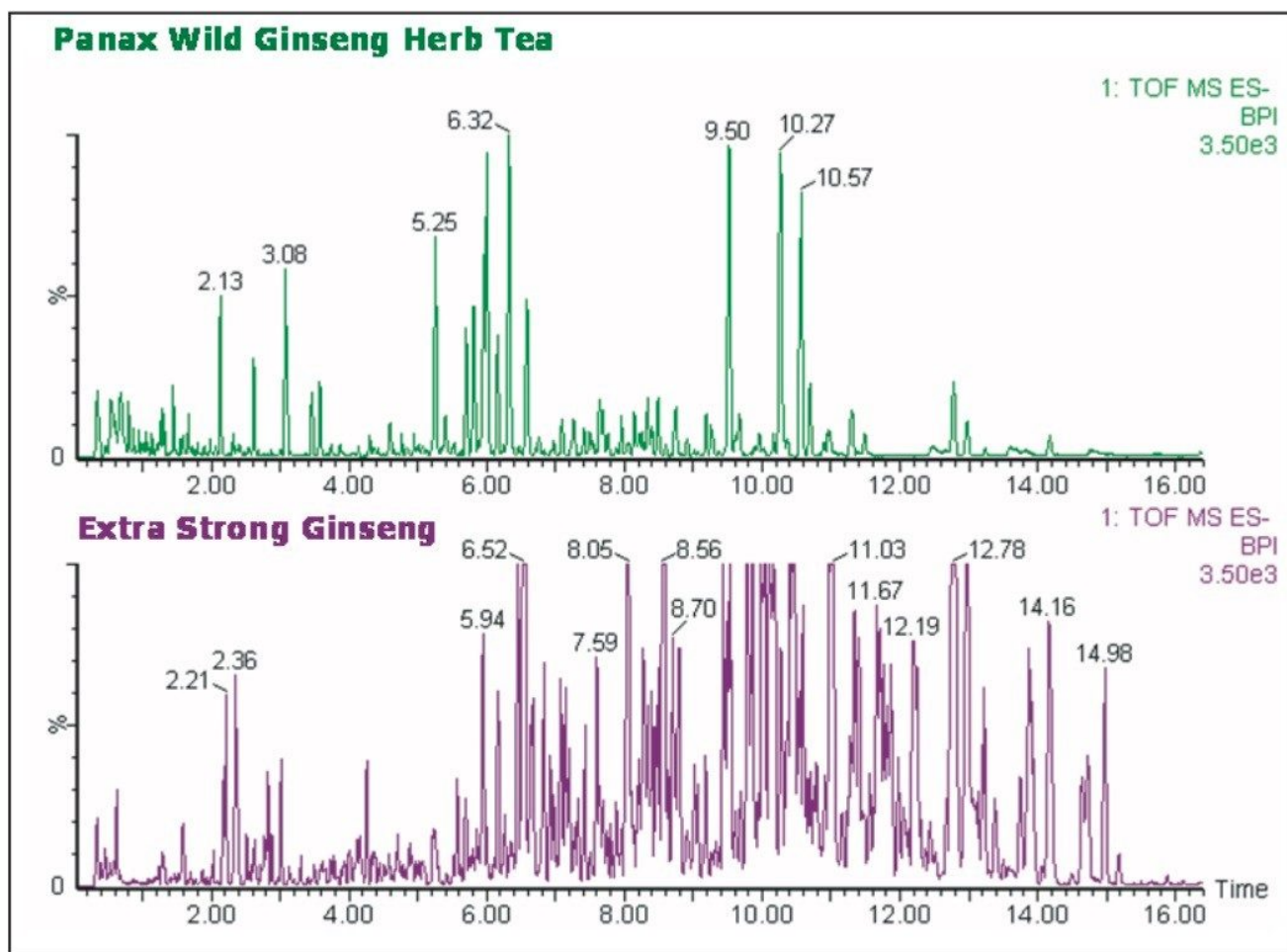


Figure 2. LC-MS base peak ion chromatograms for the two Ginseng extract samples.

The first step for multivariate statistical analysis of the LC-MS dataset was to convert the 3D LC-MS data into a 2D matrix. This critical step was accomplished by using the MarkerLynx, an Application Manager for MassLynx Software. MarkerLynx converts each data point into an Exact Mass Retention Time (EMRT) pair and tabulates the results into a 2D matrix (Figure 3).

There were 1184 EMRT pairs found here. The number of EMRT pairs detected depends on the peak detection threshold, which is a userdefined parameter.

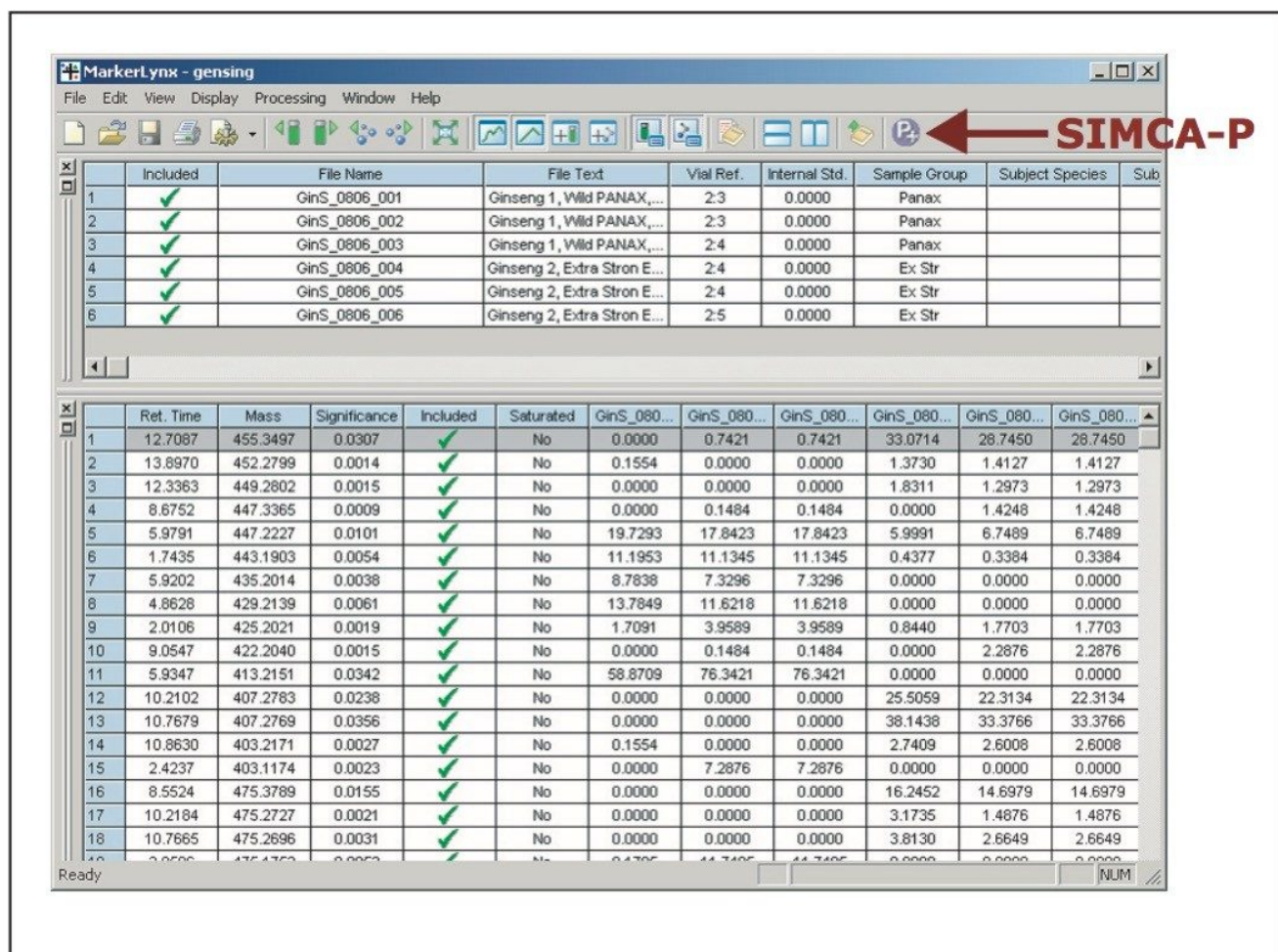


Figure 3. The MarkerLynx result window. The top window shows the list of sample injections. The bottom window shows the EMRT table.

The EMRT table can be automatically imported into SIMCA-P by simply clicking on the P+ button. The data is processed first by using principal component analysis (PCA). Then a minimum supervised statistical model, Orthogonal Partial Least Square-Data Analysis (OPLS-DA), can be applied for orthogonal data analysis. Figure 4 shows the scores plot obtained as the result of the OPLS-DA. The scores plot clearly displays the differences of the two sample groups along the x-axis as well as the differences within the same sample group along the y-axis.

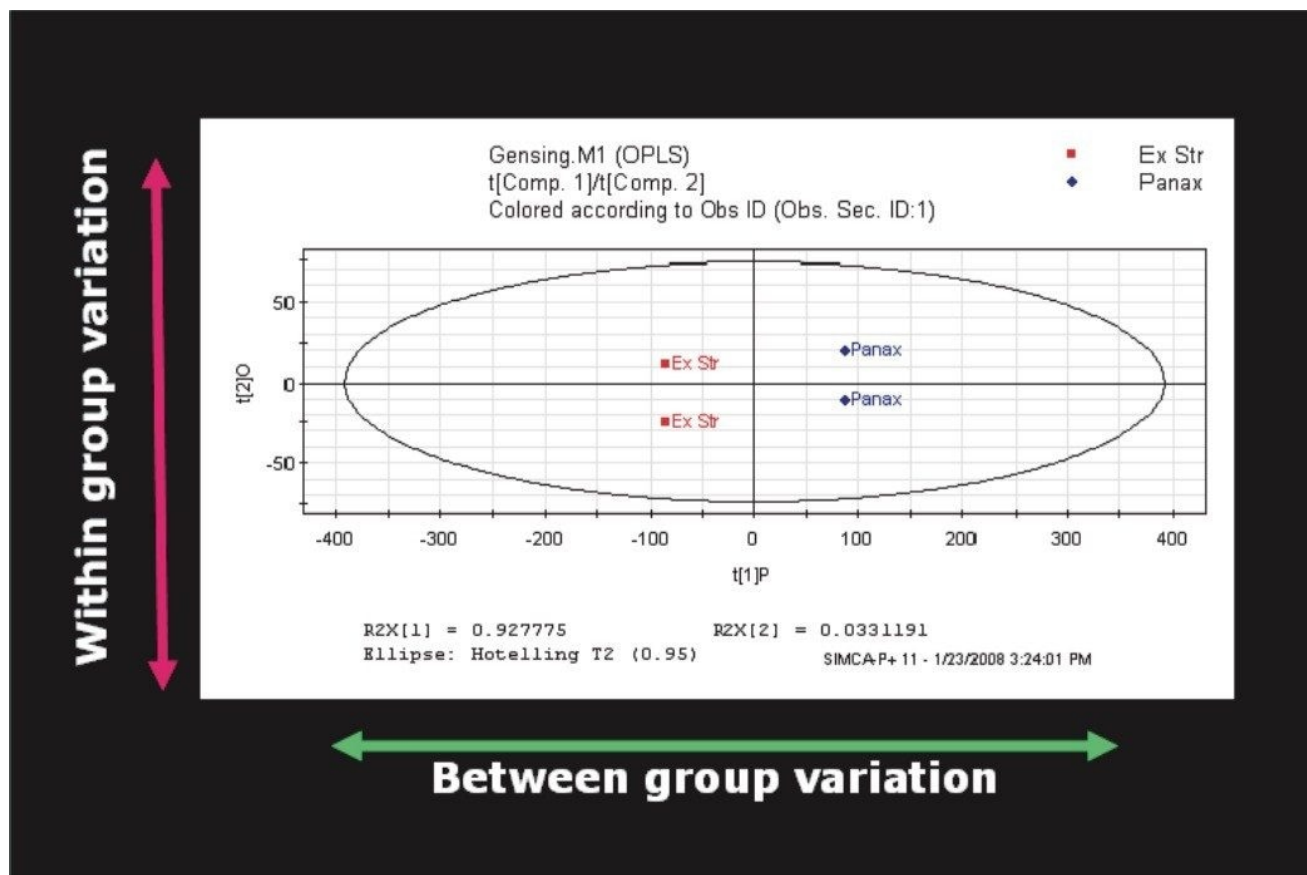


Figure 4. The scores plot, showing a clear grouping between the Extra Strong Ginseng and the Wild Panax Ginseng.

To further chemically identify the differences between the two sample groups, a scatter plot (S-plot) based on the OPLS-DA was obtained and is shown in Figure 5.

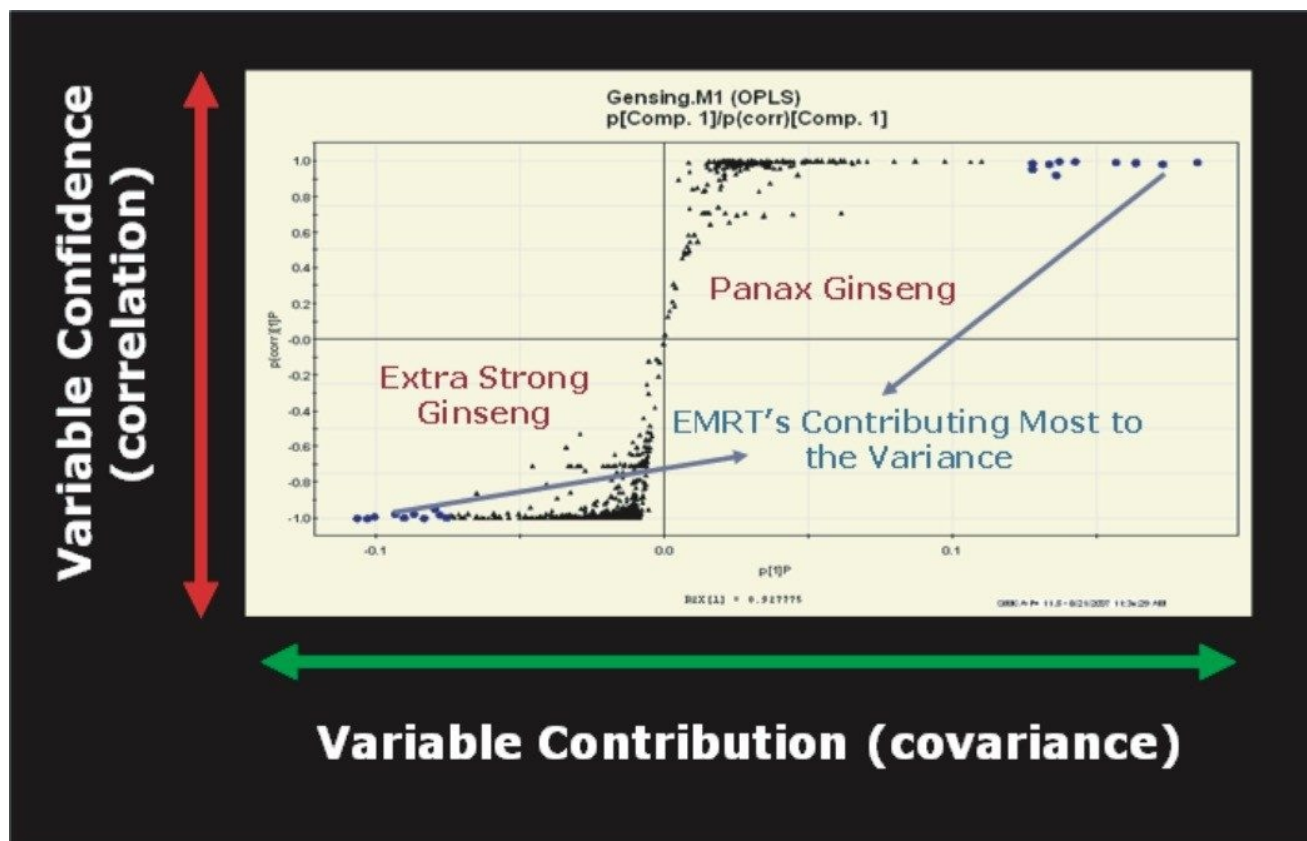


Figure 5. S-plot of the Extra Strong Ginseng vs. Wild Panax Ginseng obtained based on the OPLS-DA statistical analysis of the dataset.

In the S-plot, each point represents an EMRT pair. The x-axis shows the variable contributions. The further away a data point from 0, the more contribution it has for the variance from the sample. The y-axis shows the sample correlations within the same sample group. The further away an EMRT pair from the value 0, the better correlation it has among the injections. As a result, the EMRT pairs on both ends of the S-shaped curve represent the leading contributing ions from each sample group with the highest confidence.

For example, in Figure 5, the EMRT pairs close to upper-right corner of the S-plot are the leading contributing markers from the Wild Panax Ginseng with high confidence; the EMRT pairs close to the lower left corner of the S-plot are the leading contributing markers from the Extra Strong Ginseng with high confidence.

The leading contributing EMRT pairs can be selectively captured so that a list of top contributing markers for each sample group can be generated and saved as a text file. This text file can be later imported back into

MarkerLynx as a results table for elemental composition searches as well as database searches. Figure 6 shows the two lists of the top 10 leading EMRT pairs obtained from the S-plot for both the sample groups.

Figure 6 shows the fact that the m/z 945.5419 ion at retention time 6.54 min is the most significant marker in the Extra Strong Ginseng, and this is at the confidence level of 0.999. And the m/z 801.5021 ion at retention time of 6.33 min is the most significant marker in Wild Panax Ginseng, with the confidence level of 0.994.

In addition, the top 10 EMRT pairs are in lower MW ranges in Wild Panax Ginseng (from m/z 623 to m/z 955) compared with those found in Extra Strong Ginseng (from m/z 783 to m/z 1187). This indicates that the top 10 markers in Extra Strong Ginseng contained mostly 3–4 sugar rings, while the top 10 markers in Wild Panax Ginseng contained mostly 2–3 sugar rings.

Extra Strong Ginseng	Covariance	Correlation	Panax Wild Ginseng	Covariance	Correlation
6.54_945.5419	-0.107	-0.999	6.33_801.5021	0.185	0.994
8.62_783.4892	-0.104	-1.000	10.28_801.5020	0.173	0.984
10.40_1077.5919	-0.101	-0.990	10.58_801.5025	0.164	0.992
10.04_1107.6005	-0.094	-0.978	6.01_801.5058	0.164	0.992
11.00_945.5472	-0.091	-0.995	9.54_955.4945	0.157	0.995
10.17_1077.5924	-0.087	-0.974	5.25_817.4964	0.143	1.000
6.45_1187.5520	-0.084	-1.000	3.08_623.1978	0.137	1.000
6.50_1157.5475	-0.080	-0.948	2.13_785.2483	0.136	0.922
11.36_1047.5763	-0.078	-0.981	5.96_701.4496	0.134	0.984

Figure 6. The top 10 leading EMRTs for each sample group obtained from the OPLS-DA S-plot.

The top 10 EMRTs can also be reviewed in a bar chart format. Figure 7 shows the bar charts for the top 10 markers for Extra Strong Ginseng (7a) and Wild Panax Ginseng (7b).

The bar chart offers additional information for the markers that were already identified on the list, showing a direct comparison of the top 10 EMRTs between the two sample groups in question. In figure 7, the top 10 markers from Extra Strong Ginseng were barely detected in the Wild Panax Ginseng. While the top 10 markers from Wild Panax Ginseng were detected with very low intensities from the Extra Strong Ginseng, some were undetectable as well.

In addition, some semi-quantitative information is provided by the bar chart. The top 10 markers of Wild Panax Ginseng were detected at much higher intensities than the top 10 markers from the Extra Strong Ginseng. This is an indication that the Wild Panax Ginseng drink is a cleaner extract than the Extra Strong Ginseng drink.

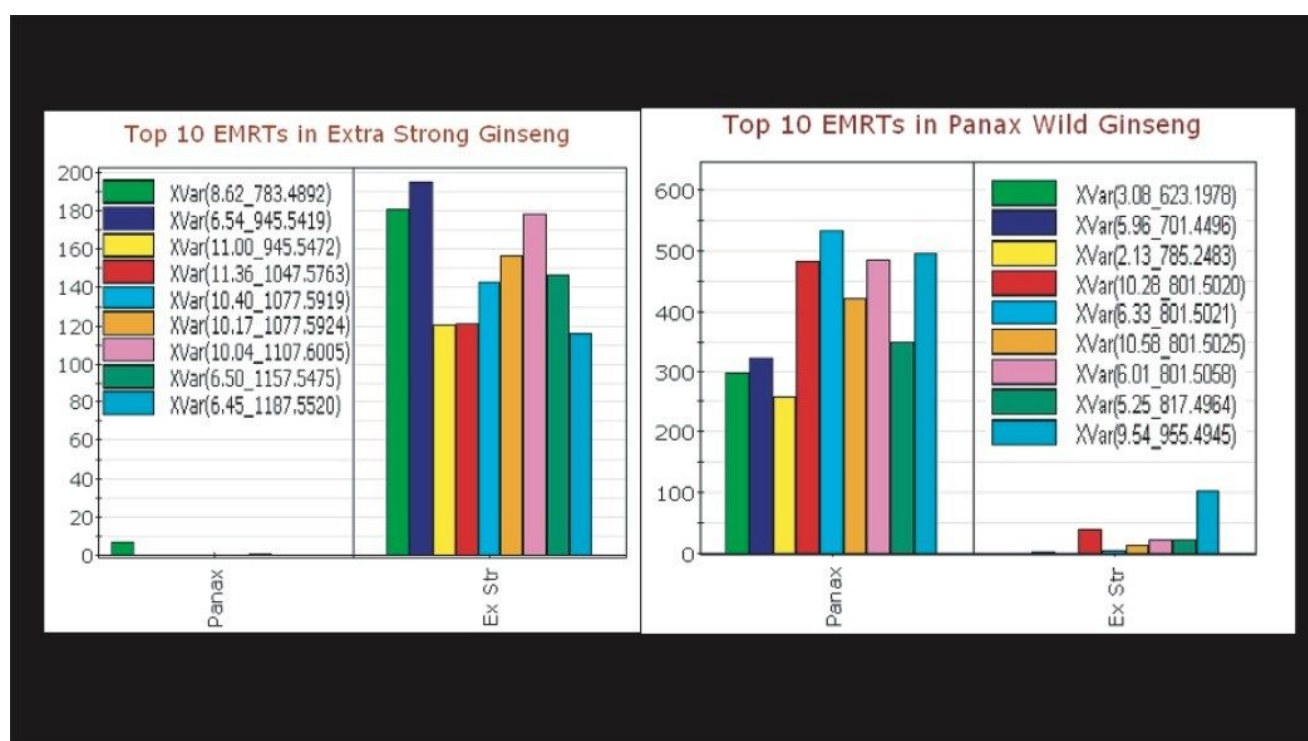


Figure 7. Bar charts of the top 10 markers for Extra Strong Ginseng (7a) Wild Panax Ginseng (7b).

As mentioned, the text files obtained from SIMCA-P can be directly imported to the MarkerLynx results table. Figure 8 shows the screen shot of the MarkerLynx result window with the two results tables filled, one for each sample group.

MarkerLynx - gensing_KY

File Edit View Display Processing Window Help

	Included	File Name	File Text	Vial Ref.	Internal Std.	Sample Group	Subject
1	✓	GinS_0806_001	Ginseng 1, Wild PANAX...	2:3	0.0000	Panax	
2	✓	GinS_0806_002	Ginseng 1, Wild PANAX...	2:3	0.0000	Panax	
3	✓	GinS_0806_003	Ginseng 1, Wild PANAX...	2:4	0.0000	Panax	
4	✓	GinS_0806_004	Ginseng 2, Extra Stron E...	2:4	0.0000	Ex Str	
5	✓	GinS_0806_005	Ginseng 2, Extra Stron E...	2:4	0.0000	Ex Str	
6	✓	GinS_0806_006	Ginseng 2, Extra Stron E...	2:5	0.0000	Ex Str	

	ID	Ret. Time	Mass	Significan...	Elemental Composition (mDa)	SIMCA1	SIMCA2
27		6.3200	655.4440	0.0337	1.9, 0.0, C36H63O10	0.0652	0.9959
28		9.6831	783.4911	0.0316	(1) 1.6, 0.1, C42H71O13	0.0616	0.9950
29		11.3107	799.4872	0.0312	(1) -3.1, 0.1, C35H75O19	0.0577	0.9955
30		2.1238	623.2181	0.0311	(1) -0.6, 0.1, C26H39O17	0.0627	0.9998
31		8.2577	817.4988	0.0306	(1) 3.9, 0.2, C42H73O15	0.0296	0.9901
32		7.9732	783.4939	0.0305	(1) -1.4, 0.1, C35H75O18	0.0584	0.9977
33		2.6151	623.2183	0.0297	(1) -0.4, 0.0, C26H39O17	0.0614	0.9996
34		7.2650	799.4863	0.0292	(1) 1.9, 0.7, C42H71O14	0.0576	0.9990
35		11.5268	783.4906	0.0279	(1) -2.4, 0.7, C60H63	0.0266	0.9696
36		3.0750	461.1667	0.0273	0.8, 0.0, C20H29O12	0.0589	1.0000
37		6.5738	815.4813	0.0262	(1) -3.9, 0.3, C35H75O20	0.0492	0.9999
38		8.5000	971.4880	0.0248	(1) 2.8, 0.6, C48H75O20	0.0528	1.0000
39		6.0030	493.3894	0.0242	0.1, 0.0, C30H53O5	0.0546	0.9981
40		9.2750	799.4792	0.0238	0.7, 0.0, C49H67O9	0.0527	0.9999
41		7.7624	969.4721	0.0231	(1) 2.6, 0.3, C48H73O20	0.0324	0.9958
42		7.4174	799.4871	0.0227	(1) -0.8, 0.5, C60H63O	0.0447	0.9964
43		6.0073	847.5090	0.0226	3.5, 0.0, C43H75O16	0.0530	0.9982
44		10.2831	639.4478	0.0211	0.6, 0.0, C36H63O9	0.0502	0.9999
45		6.3220	847.5088	0.0208	3.3, 0.0, C43H75O16	0.0511	0.9967

Ready CAP

M/Z 971.4880
C₄₈H₇₆O₂₀

Figure 8. The screen shot of the MarkerLynx results table with the EMRTs imported .txt file obtained from SIMCA-P S-plot.

From the MarkerLynx results table, the exact mass reported for each EMRT pair can be searched for elemental composition. This information can be used for further querying of existing databases to find putative chemical structures (if the marker found resides in the database).

As an example, we chose a marker from Panax Ginseng with m/z 971.4880 and elemental composition of C₄₈H₇₆O₂₀ to search a database available from the public domain, Chempidder. One of the possible hits is shown in Figure 9.

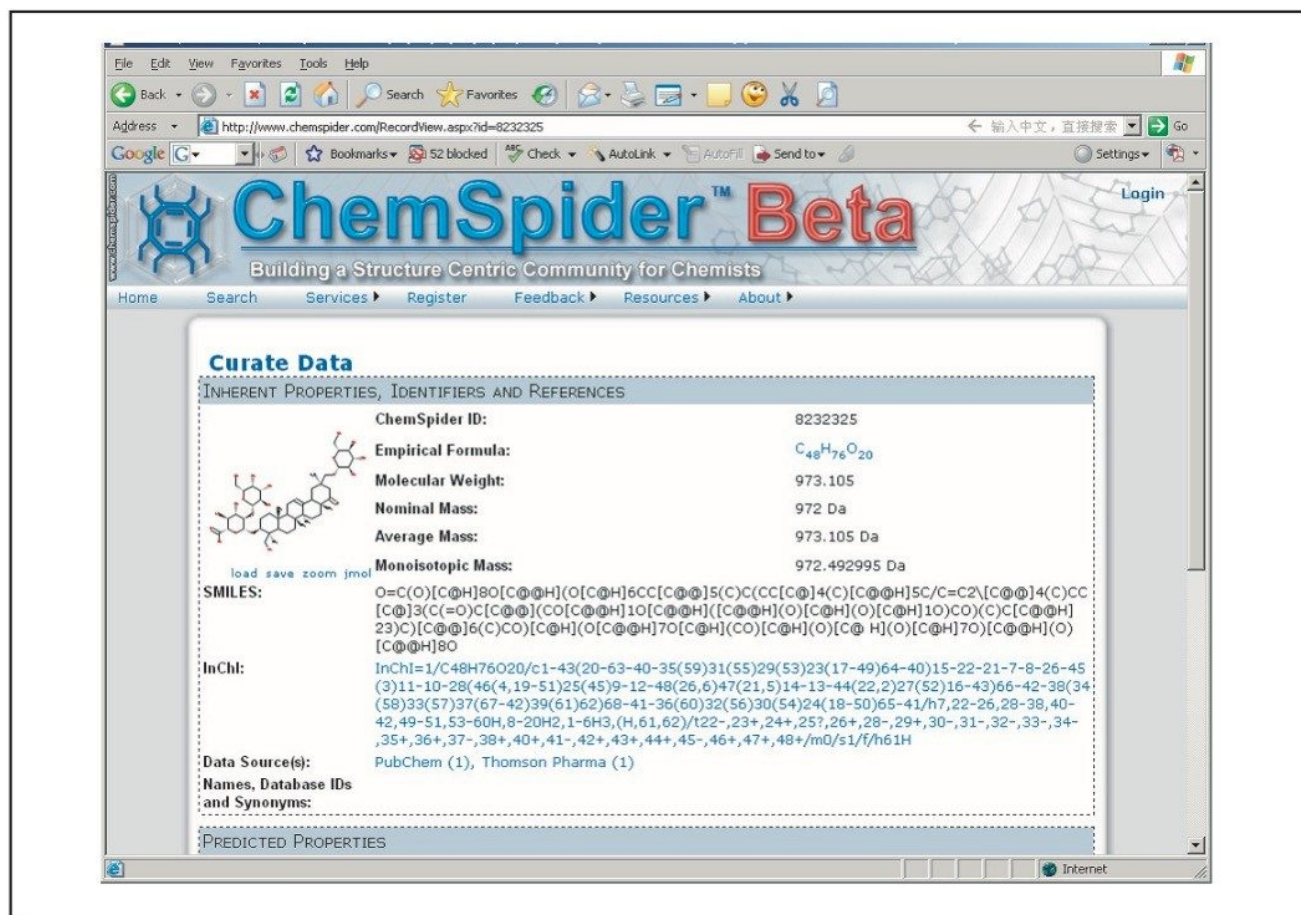


Figure 9. The possible hit found from the Chemspider database for marker m/z 971.4880.

From this information, it is easy to go back to the LC-MS raw data and confirm the structure found using the fragment ions obtained from the TOF MS^E data.¹

Conclusion

This application note demonstrates generic intelligent workflow for Traditional Herbal Medicine (THM) sample profiling. This approach is very effective for general comparison of extremely complex samples.

By using the ACQUITY UPLC-SYNAPT HDMS systems with TOF MS for analysis, raw data with exact mass

measurements are first collected. The multivariate statistical analysis can be performed for the dataset after they are converted into a 2D matrix as an EMRT pair. The top contributing ions for each sample can be easily obtained from the OPLS-DA S-plot generated in SIMCA-P. Results can be imported back into the results table in MarkerLynx. A database query can then be performed to obtain elemental composition, as well as chemical structure if the marker is a characterized compound.

This entire approach is easy, fast, and generic. It can be easily adapted for profiling various types of THM samples. As a result, significant resource savings can be accomplished with maximum information obtained.

References

1. An Intelligent Workflow for Traditional Herbal Medicine: Compound Identification by UPLC-TOF MS. Yu K, Castro-Perez J, Shockcor J. Waters Application Note. 2008; [720002486EN](#) <<https://www.waters.com/webassets/cms/library/docs/720002486en.pdf>> .

Featured Products

[ACQUITY UPLC System <https://www.waters.com/514207>](https://www.waters.com/514207)

[Progenesis QI Software <https://www.waters.com/134790655>](https://www.waters.com/134790655)

720002541, March 2008

© 2022 Waters Corporation. All Rights Reserved.

[Terms of Use](#)

[Privacy](#)

[Trademarks](#)

[Sitemap](#)

[Careers](#)

[Cookies](#)

[Preferencias de cookies](#)