[应用纪要]

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

多元数学统计方法分析传统草药，使用 UPLC 超高效液相色谱/TOF-MS 飞行时间质谱比较不同样品种类

Kate Yu, Jose Castro-Perez, 和 John Shockcor
沃特世公司，米尔福德，马萨诸塞州，美国

前言

传统草药（THM）或传统中药（TCM）样品的分析研究是非常具有挑战性的，直接原因是样品的重现性差。植物提取物的成分会因产地，采收季节以及提取方法的不同而发生显著变化。即使提取物是来自同一株植物的提取物或来自相同名称的两株植物，其成分也不尽相同。

此外，为了有效的对中药进行质量控制，非常有必要对中草药进行分析比较。中草药样品分析对于传统草药的生理作用机理的研究也是非常关键的。

我们开发了一套简便快速且易于通用的传统中草药分析流程的(图 1)。该分析流程利用了沃特世（Waters®）UPLC® 超高效液相色谱的技术优势，即高分辨，高灵敏度和快速分离，并结合了 SYNAPT™ HDMS™ 质谱系统的飞行时间质谱仪(TOF MS)精确质量数测定的功能。该工作流程能够应用于化合物鉴定或样品解析。

传统中草药中的化合物鉴定在我们已在另一篇应用纪要中讨论过。[1] 本文将演示如何利用该分析流程借助多元数学统计方法进行样品数据的解析。结果表明，样品的比较可以在几个小时内完成并获得完整的样品信息。这显著地缩短了传统草药样品的分析时间和节省了人力。

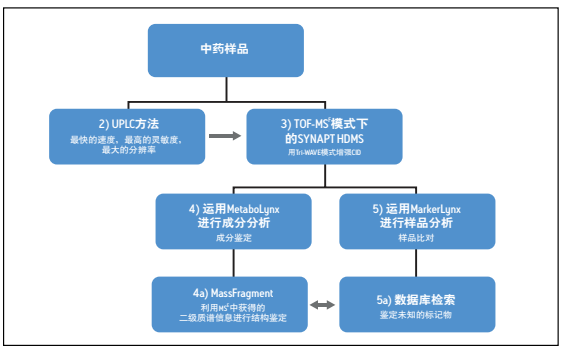

图 1. 传统草药分析的工作流程。

实验方法

本实验的样品来自于两种人参提取物口服液。

- 样品 1 是人参精口服液（产自中国，JV Trading Ltd. 公司销售，纽约，纽约州）。
- 样品 2 是青春宝口服液（产自中国，Overseas Factor Corporation 公司销售，旧金山，加利福尼亚州）。

每个样品在进样前先过滤。

液相条件

液相系统:	沃特世 ACQUITY UPLC® 超高效液相色谱系统		
色谱柱:	ACQUITY UPLC 超高效液相色谱 HSS T3 色谱柱 2.1 x 100 mm, 1.7 µm, 65 °C		
流速:	600 µL/min		
流动相 A:	水+ 0.1% 甲酸		
流动相 B:	甲醇		
梯度:	时间	组成	曲线
	0 min	95% A	
	10 min	30% A	Curve 6
	17 min	0% A	Curve 6
	20 min	95% A	Curve 1

质谱条件

质谱系统:	沃特世 SYNAPT HDMS 质谱系统
离子化模式:	电喷雾
毛细管电压:	3000 V
锥孔电压:	35 V
除溶剂温度:	450 °C
除溶剂气体:	800 L/Hr
离子源温度:	120 °C
采集范围:	50 to 1500 m/z
碰撞气体:	氩气

数据处理

化合物筛选和分析: MarkerLynx™ 应用管理软件

多元数学统计分析: SIMPCA-P

结果

为保证数学统计结果的可靠性和重要性，每个样品至少重复进样三次。为获得每个样品的所有信息，有必要对它们在正负离子模式下进行LC/MS分析。本实验中，每种样品重复进样六次:三次电喷雾正离子模式分析和三次电喷雾负离子模式分析。出于演示目的，本文只讨论了负离子模式下的结果。

图 2 显示两种人参提取物口服液基峰离子色谱图的比较。由图可以看出人参精口服液含成份远多于青春宝并且浓度更高。由于两个样品成份都很复杂，有必要利用多元数学统计工具对两个样品做进一步的分析。

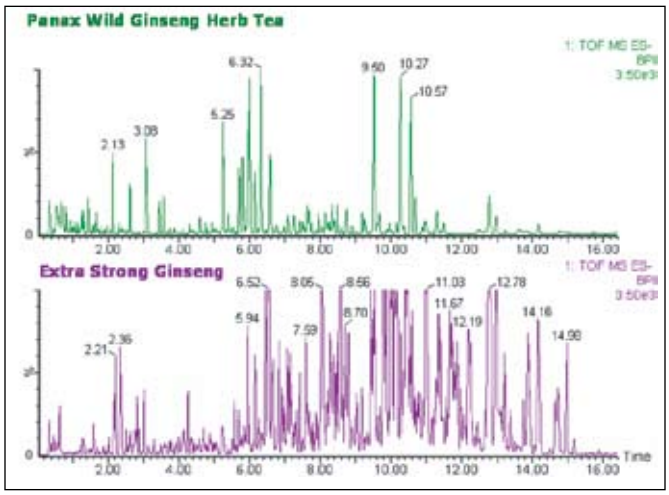

图 2. 两种人参提取物样品的 LC/MS 液相色谱/质谱基峰离子色谱图。

使用多元数学统计方法对 LC/MS 数据进行分析的第一步是将三维 LC/MS 数据转换成二维矩阵。这一关键步骤由 MassLynx™ 操作软件中的 MarkerLynx 完成。MarkerLynx 将每一个数据点转换成精确质量保留时间 (EMRT) 数据对，并以二维矩阵型式将结果列出 (图 3)。

本实验共得到了 1184 个精确质量保留时间 (EMRT) 数据对 。可检测到 EMRT 数据对的数量取决于色谱峰检测限的设定，该参数可由分析人员设定。

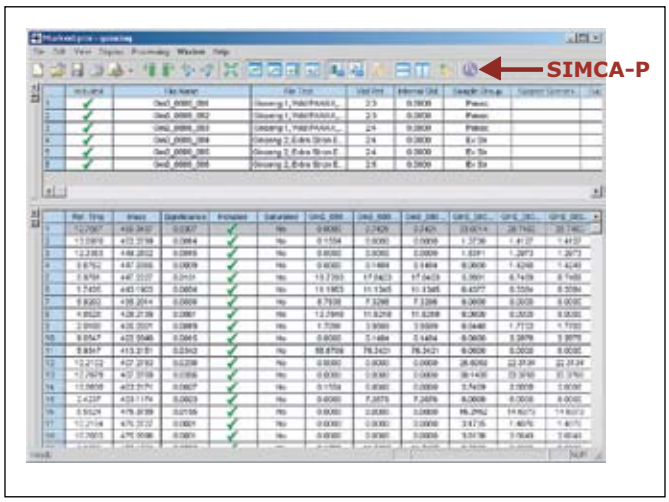

图 3. MarkerLynx 结果显示窗口。窗口上部为样品进样列表。窗口下部为精确质量于保留时间数据对列表。

从 MarkerLynx 报告界面上，仅需点击 P+ 按钮，EMRT 数据对列表就可以被自动导入到 SIMCA-P 中。首先利用主成分分析 (PCA) 法对对数据进行处理。之后利用无监督统计学模型，结合正交偏最小二乘法进行两维数据分析 (OPLS-DA)。图 4 列出正交偏最小二乘法数据分析的分值结果。该图清晰地展示了两个样品组在 X 轴和 Y 轴方向的差别。

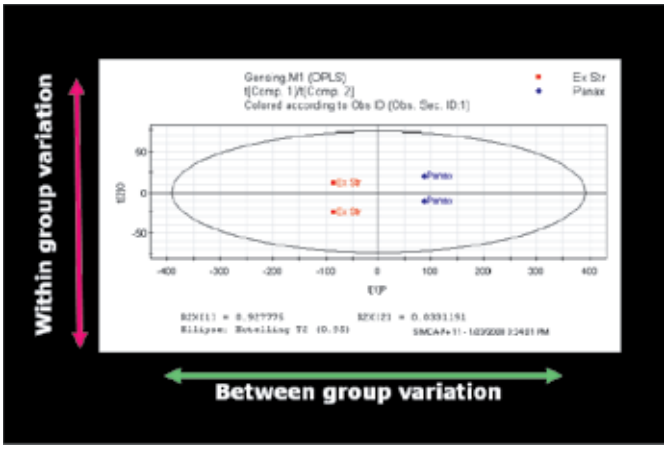

图 4. 数值图表示人参精口服液和青春宝口服液明显的分组情况。

为进一步鉴定两组样品的化学组成上的差异性，正交偏最小二乘法得到的数据分析结果散点图如图 5 所示。

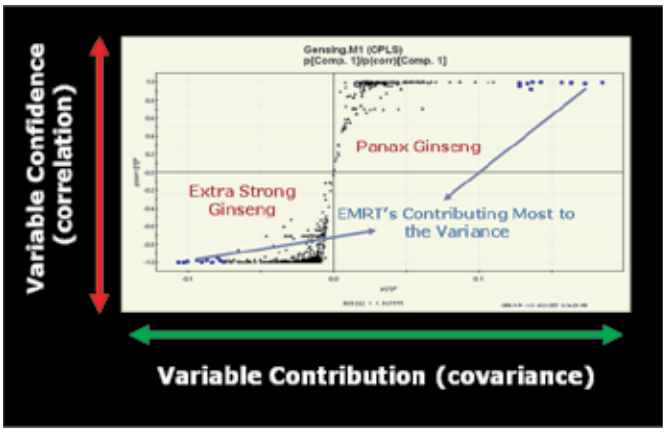

图 5. 基于正交偏最小二乘法获得的人参精口服液和青春宝口服液数据分析结果散点图。

在散点图中，每个点代表一个精确质量保留时间数据对。X轴表示可变量。一个数据点距离 0 越远，该点对样品差异的贡献越大。Y 轴表示在同一样品组中的样品间的相关性。精确质量与保留时间数据对距离 0 值越远，进样间的相关性越好。因此，在 S 型曲线两端的 EMRT 数据对代表了来自每个样品组的可信度最高的特征离子。

例如，图 5 中，接近 S 图右上角的 EMRT 数据对为来自青春宝口服液可信度最强的特征标记物，接近 S 图左下角的 EMRT 数据对为来自人参精口服液可信度最强的特征标记物。

这些特征的 EMRT 数据对可以被选择性地捕获，并获得每组样品中特征标记物列表，并以 TXT 文件保存下来。这个 TXT 文件可被输回 MarkerLynx，产生一个结果列表，从而用于元素组成搜索以及数据库搜索。图 6 显示了从两组样品 S 图中获得的十个特征的精确质量与保留时间数据对列表。

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

图 6. 利用正交偏最小二乘法从两个样品数据分析散点图中获得的最高贡献的十个精确质量保留时间数据对列表。

图 6 表明保留时间为 6.45 分钟质荷比为 945.5419 离子是人参精样品中最显著的标记物，可信度达 0.999。保留时间为 6.33 分钟质荷比为 801.5021 的离子是青春宝样品中最显著的标记物，可信度达 0.994。

此外，相比人参精样品(从质荷比 783 到质荷比1187)，青春宝样品中最特征的十个 EMRT 数据对在较低的分子量范围内(从质荷比 623 到质荷比 955)。这说明人参精样品的十个特征的标记物中的大多数含有三至四个糖环，而青春宝样品中最特征的十个标记物含有二至三个糖环。

差异性最大的十个 EMRT 数据对也可以用棒状图格式进行查看。图 7 列出人参精 (7a) 和青春宝 (7b) 十个差异性最大的标记物的棒状图。

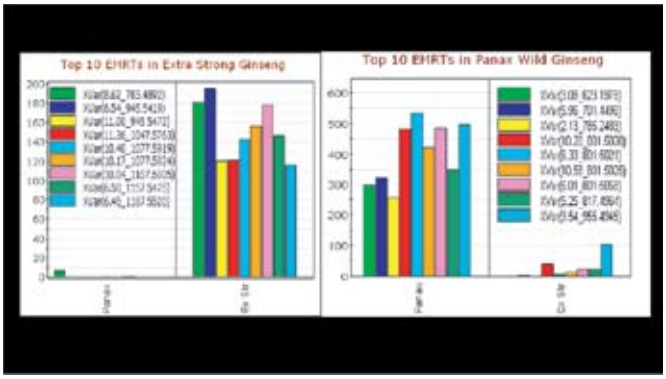

图 7. 人参精 (7a)和青春宝(7b)十个差异性最大的标记物的棒状图。

棒状图提供了列表中已经鉴定的标记物的额外信息，显示被研究的两个样品组十个差异性最大的 EMRT 数据对的直接比较结果。在图 7 中，人参精样品的十个特征标记物在青春宝样品中几乎没有被检测到。而来自青春宝样品的十个特征标记物在人参精样品中被检测到具有很低的强度，有些也未能检测到。

此外，棒状图也提供了一些半定量的信息。来自青春宝样品的十个最大标记物比在人参精样品中检测到的强度高。表明青春宝口服液是比人参精口服液更纯的提取物。

如上所述，从SIMCA-P得到的文本文档可以直接导入MarkerLynx结果列表中。图 8 显示填入两组结果的 MarkerLynx 结果窗口界面，每个表格代表一组。

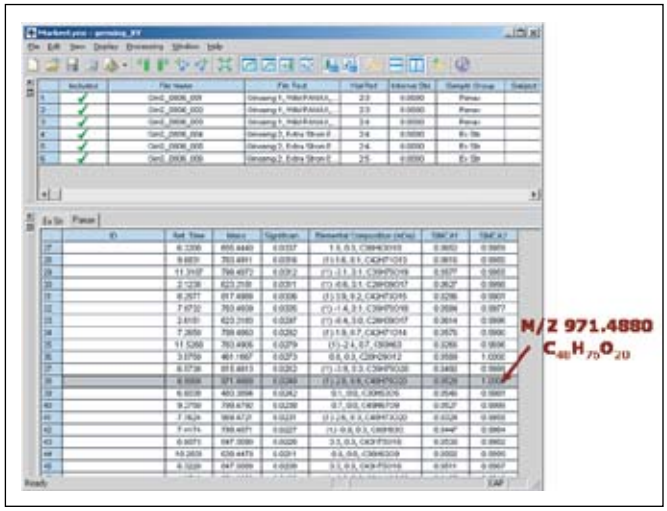

图 8. 导入精确质量与保留时间数据对的 MarkerLynx 结果显示窗口界面，文本文档从 SIMCA-P 散点图获得。

从 MarkerLynx 结果表格中，可以对每一个 EMRT 数据对报告中的精确质量进行元素组成分析检索。此信息可进一步用于作现有数据库搜查，寻找推断的该成分的化学结构(如果数据库中存在该种标记物)。

举例来说，我们从青春宝样品中选择一个质荷比为 971.4880 的标记物，其元素组成为 C48H76O20，对公共平台数据库，Chemspider 进行检索。其中一个可能性如图 9 所示。

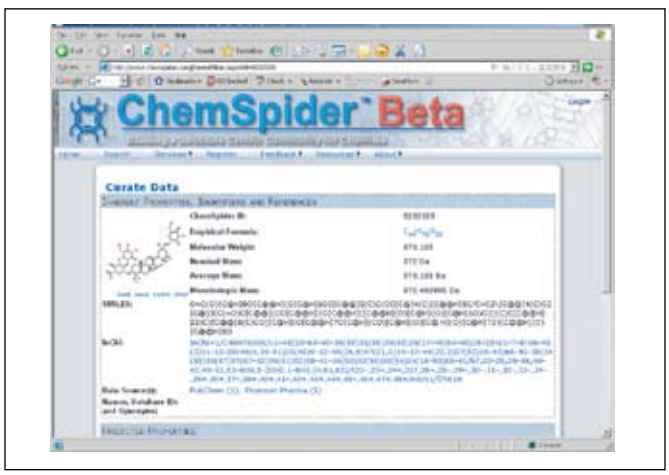

图 9. Chemspider 数据库中检索的到的质荷比 971.4880 的可能结构。

从该信息很容易返回到液相色谱/质谱 LC/MS 原始数据，利用飞行时间 TOF MSE 数据[1]的碎片离子来确认推导的结构的准确性。

结论

本应用文集演示一种通用智能化的传统中草药样品分析的工作流程。相对于传统的分析方法，当前这种方法对于相当复杂样品的分析非常有效。

通过 UPLC® 超高效液相色谱/SYNAPT™ HDMS™ 质谱系统的进行飞行时间质谱分析，首先采集含有精确质量测定的原始数据。当将这些数据作为精确质量保留时间数据对转成二维矩阵形式，多元数学统计分析方法即可对这套数据进行分析。每个样品的最特征的离子可以从 SIMCA-P 的正交偏最小二乘法数据分析散点图中获得。结果可以导回 Markerlynx 的结果列表中。如果标记物是已经解析出的化合物，可利用数据库检索其元素组成及化学结构。

整套分析方法简便，快速适用性强。它可以很方便地应用到不同类型的传统中草药样品分析之中。因此，在显著节省资源的同时获得最大信息量。

参考文献

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.

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Waters, ACQUITY UPLC, 和 UPLC 是沃特世公司的注册商标。SYNAPT, HDMS, MassLynx, MarkerLynx, 和 The Science of What's Possible 是沃特世公司的商标。所有其它商标属于各自的所有者。

©2008 沃特世公司 中国印刷
2008 年 6 月 720002541ZH

沃特斯中国有限公司
沃特世科技（上海）有限公司

北京：010 - 8586 8899
上海：021 - 6879 5888
广州：020 - 8626 6678
香港：852 - 2964 1800

免费售后服务热线：800 (400) 820 2676
www.waters.com www.waterschina.com