

文章编号: 1006-6144(2009)01-0041-06

中药指纹图谱共有峰的自动识别

冯启余, 曹玉华*, 龚立冬

(江南大学化学与材料工程学院, 江苏无锡 214122)

摘要: 依托时间匹配、光谱信息辅助, 利用自编的 VB 程序, 建立并比较了时间匹配和时间匹配参考光谱信息两种识别中药指纹图谱共有峰的方法, 并成功地应用于肉苁蓉的高效液相色谱-二极管阵列检测器(HPLC-DAD)法指纹图谱的建立。结果表明, 只需要人为设置色谱峰大小下限、保留值相对误差上限和最大吸收波长差值上限三个参数, VB 程序编制的指纹分析系统就能快速对所有样本进行共有峰识别和模糊数学处理, 实现了无人工干预下的中药指纹图谱共有峰的自动识别, 指纹图谱的建立过程客观, 并且节省大量时间, 其中参考光谱法识别结果更准确。

关键词: 中药指纹图谱; 共有峰识别; 肉苁蓉; 高效液相色谱-二极管阵列检测器

中图分类号: O657.7⁺2 **文献标识码:** A

中药是中华民族的瑰宝, 现代化是中药走向世界、实现国际化的前提, 其制剂的质量水平控制是现代化的中心环节^[1]。以指纹图谱^[2,3]规范中药制剂的质量, 被评价为中药质量控制的里程碑^[4-6]。国家食品药品监督管理局 2002 年发布了《中药注射剂色谱指纹图谱实验研究技术指南(试行)》^[7]。色谱法是研究指纹图谱的首选方法, 高效液相色谱(HPLC)因具有灵敏、重现、适用对象广泛、分离效率高等特点而被广泛应用^[8]。

峰匹配是指纹分析的关键, 保留时间或相对保留值常常作为色谱指纹图谱的共有峰识别标准^[9]。但是, 由于中药体系的复杂性和 HPLC 仪器系统的微小波动, 同一组分在各色谱图中的保留时间不尽相同。通常采用添加内标物或者人工选定标定峰来校正保留时间^[10]。根据选定标定峰的数目有一点、两点或者多点校正方法。一点法对非线性推移校正结果较差, 两点法可以部分校正非线性推移, 多点法效果最好, 但是加大了操作者的难度^[11]。陈闽军等^[10]用遗传算法进行色谱图匹配, 实现了自动优化标定, 但需要在色谱图中分区间定义局部特征, 有可能导致区间邻接位置的同一色谱峰被划分入不同的区间, 使自动匹配结果有时与观察特征不一致, 须通过手动校正进行调整。朱臻宇等^[12]提出一种全排序模板匹配算法, 克服了原算法中的某些错配、漏配等缺陷。但是, 全排序模板是所有样本的总和, 数据量大, 不利于大批量样本和峰数目很多的样本的处理。当样本之间差异较大时, 其同一物质的色谱峰应当彼此靠近的分析前提也难以保证。倪力军等^[13]选取前后两个参数点, 采用图论中的最短路径算法寻找最佳匹配峰组, 最大程度地匹配可能的色谱峰并极少出现错配、漏配峰组, 无需手动校正。因此以保留时间为基础, 结合其它色谱信息综合进行峰匹配成为必然。王龙星等^[14]提出了在保留时间的一定范围内参考峰高的算法, 能较好地定量评价指纹图谱间的相似性。但是峰高的定性作用很不可靠, 当样品间的组分差异比较大或者大小相似的峰密集出现时, 峰高的参考意义不大。二极管阵列检测器(DAD)灵敏度高、重现性好, 能得到光谱信息^[15], 可以反映检测对象的部分结构信息, 有利于复杂体系的峰识别^[16-19]。李博岩等^[16]提出了光谱相关色谱的概念, 并且分析处理了一些中药色谱指纹图谱, 结果表明该方法实用、可靠, 仍有待进一步发展和完善。但在具体应用中还要借助人工干预, 而非自动地全图谱处理。

收稿日期: 2008-04-16 修回日期: 2008-06-15

基金项目: 国家自然科学基金(No. 20775028)

*通讯联系人: 曹玉华, 女, 教授, 研究方向为中药分析。

本文依托时间匹配、光谱信息(最大吸收波长)辅助,利用 VB6.0 自编的程序,实现了无人工干预下的中药指纹图谱共有峰的识别及相似度分析,有利于指纹图谱的客观化和标准化。

1 算法设计

1.1 时间匹配识别共有峰

本文采用已知活性组分的峰作为参考峰,综合了内标法的优势,优先保证活性物的匹配有利于反映主要信息。两个不同样本中,按照相对保留时间最接近的峰认为是共有峰的原则,设计时间匹配算法,具体算法如下:取两个样本 P1、P2,依次取 P1 中的峰 i,与 P2 中未匹配的相对保留值最小的峰 j 比较。如果差值不大于预设值 N_t (其取值根据实际情况而定),认为有可能是共有峰,记录并取 P2 样本中的下一个峰比较;如果差值大于 N_t ,并且 $P2_j$ 的相对保留值比 $P1_i$ 小,继续比较,若 $P2_j$ 的相对保留值比 $P1_i$ 大,说明 P2 中与 $P1_i$ 有可能匹配的峰已经找完。取出其中与 $P1_i$ 最接近的峰 $P2_m$,如与 $P1$ 样本的下一个峰 $i+1$ 的相对保留值差大于与 i 峰的差值,说明 i 峰与 m 峰匹配,取两峰各项参数的均值记为准共有图谱的一个峰;反之,说明 i 峰与 m 峰不是最佳匹配,如 $m-1$ 峰是可能峰,则 i 峰与 $m-1$ 匹配,否则,认为 P2 中没有峰与之匹配。完成以上计算,继续取 P1 样本中的下一个峰按照上述方法识别,直至 P1 样本峰全部识别结束,得到一个准共有图谱;再以准共有图谱为标准(新 P1)识别下一个样本(新 P2),直至所有样本识别完成,得到共有图谱作为指纹图谱。其流程见图 1。

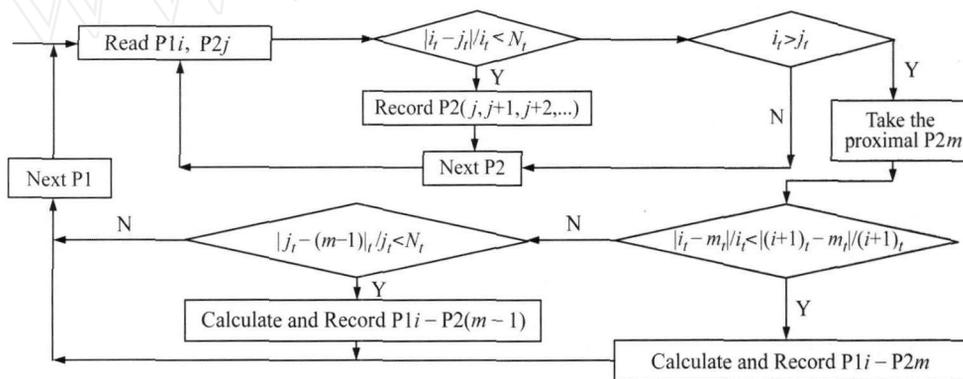


图 1 时间匹配算法流程图
Fig. 1 Schematic diagram of time matching arithmetic

1.2 参考光谱识别共有峰

在出峰比较密集的情况下,仅用相对保留值作为定性依据,结果不可靠。二极管阵列检测器可以在 200~400 nm 波长范围内得到任意保留时间色谱峰的吸收光谱,同一组分在两个不同色谱图中,虽对应的保留时间有些差异,但它们的光谱相同或相似^[16],可以利用各色谱峰的吸收光谱进行共有峰辅助识别。

依次取样本 PP1(或者准共有样本)中的峰 i,在样本 PP2 中找出所有与 i 的相对保留值相差不超过 N_t 的可能峰。选出其中最大吸收波长 λ_{max} 与 i 峰最接近的 m 峰,如果波长差值大于预设值 N_l ,认为没有匹配峰,否则反算 m 峰与 $i+1$ 峰的相对保留值和 λ_{max} 的差值。若大于与 i 峰的差值则判定 i 峰与 m 峰匹配;反之,如果 $m-1$ 峰是可能峰则判定 i 峰与 $m-1$ 峰匹配,否则认为 P2 中无匹配峰。相匹配的峰各项参数的算术平均值作为准共有样本的一个峰,所有样本的准共有样本作为指纹图谱。具体算法如图 2。

2 实验部分

2.1 仪器、试剂及样品

Waters 1525 型高效液相色谱仪(美国),包括二极管阵列检测器(Waters 2996)。借助 Empower 软件进行数据采集及处理。BRANSON SB3200-T 型超声波振荡器(上海)。旋转蒸发仪(河南巩义英裕高科仪器厂)。

甲醇(色谱纯),三氟乙酸(分析纯),超纯水,松果菊苷对照品由北京大学药学院提供。

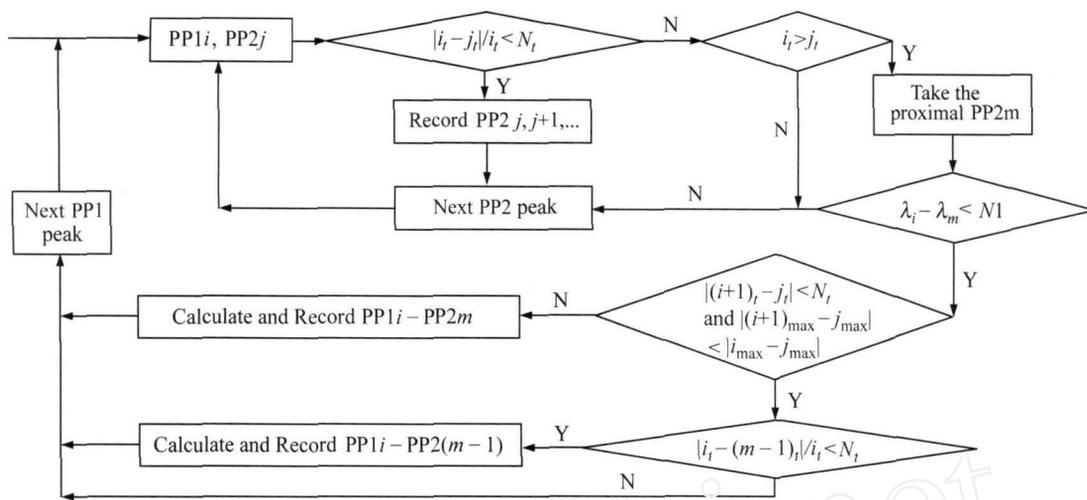


图 2 参考光谱算法流程图
Fig. 2 Schematic diagram of time spectrum matching arithmetic

实验使用 12 个不同来源的肉苁蓉 *Cistanche*^[20-22] 样品。具体信息如表 1。

表 1 样品信息表

Table 1 The information of samples

No.	Name	Cultivating area	No.	Name	Cultivating area
1	<i>Cistanche. deserticola.</i>	Neimenggu	7	<i>Cistanche. tubelosa</i>	Xingjiang
2	<i>Cistanche. lanzhouensis</i>	Lanzhou	8	<i>Cistanche. deserticola</i>	Alashan
3	<i>Cistanche. deserticola</i>	Neimenggu gudengkou	9	<i>Cistanche. deserticola</i>	Gansu yumen
4	<i>Cistanche. deserticola</i>	Ningxia	10	<i>Cistanche. salsa</i>	Qinghai
5	<i>Boschiakia. rossia</i>	Qinghai	11	<i>Cistanche. deserticola</i>	Yinchuan
6	<i>Cistanche. tubelosa</i>	Xinjiang	12	<i>Cistanche. deserticola</i>	Neimenggu

2.2 实验条件

色谱柱为 Nucleodur C₁₈ (250 × 3.9 mm i. d. , 5 μm) 柱; 流速: 0.6 mL/min。梯度洗脱程序: 0 11 min, 90% A, 10% B; 11 12 min, 90% 88% A, 10% 12% B; 12 16 min, 88% 85% A, 12% 15% B; 16 50 min, 85% 70% A, 15% 30% B; 50 115 min, 70% 55% A, 30% 45% B; 115 120 min, 55% 0% A, 45% 100% B。其中, A 相是水(含 0.1% 三氟乙酸), B 相为甲醇。二极管阵列检测器检测波长范围: 210 ~ 400 nm。进样体积: 10 μL。

2.3 实验方法

准确称取 4.0 g 肉苁蓉干燥粉末, 加入 100 mL 甲醇, 超声提取 60 min, 过滤; 残渣再分别加入 60 mL、40 mL 甲醇各提取 30 min, 合并滤液, 浓缩至 25 mL, 定容。

样品提取液经 0.22 μm 滤膜过滤, 按照 2.2 色谱条件依次测定 12 个肉苁蓉样品, 得到各样品色谱重叠图, 见图 3(从前向后依次是 1[#] ~ 12[#])。提取面积百分比大于 0.35% 的峰作为有效峰组成相应样本, 进行共有峰识别。

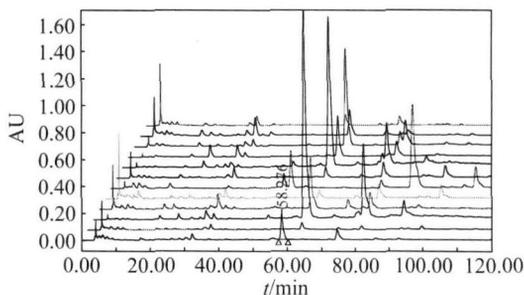


图 3 肉苁蓉样品的 HPLC-DAD 重叠图(从前向后依次是 1[#] ~ 12[#])

Fig. 3 Overlap chromatograms of *Cistanche* samples

3 结果与讨论

3.1 重现性和稳定性

方法的重现性通常用标准溶液来进行考察, 而试样

的稳定性以一个或几个特征峰来衡量,这种方式不能反映整个图谱的变化情况以及方法的重现性和稳定性。本文将所得样本,按照模糊数学方法计算相似度,则可进行全面评价。下面分别以 59 min 左右的峰(经过光谱和标准加入法鉴定为松果菊苷)进行单峰评价,以夹角余弦法计算相似度进行总体评价。

重现性试验以蒙特肉苁蓉(4[#])为试样,连续三天各测一次,在同一天测三次,检验方法的重现性。稳定性试验:将蒙特肉苁蓉提取液,每 3 h 测一次,连续进样三次,检验样品稳定性。结果见表 2。单峰评价(峰面积)与总体评价的结果大体一致,均认为方法重现性良好,样品稳定性较差,但是前者认为日内明显优于日间重现性,后者则认为重现性一致。可见,单峰评价具有片面性,采用总体评价更合理。

表 2 重现性、稳定性检验结果
Table 2 Repeatability and stability

		Holistic evaluation Similarity (vectorial angle cosine method)	Single peak evaluation	
			RSD (%) (Peak area)	RSD (%) (Relative retention times)
Repeatability	Intra-day	0.991	0.78	0.40
	Inter-day	0.991	1.8	0.82
Stability		0.980	6.8	0.48

3.2 程序可靠性检验

对同一样本进行共有峰识别,两种方法均把原样本作为准共有样本;将 12 个样本以不同顺序进行识别,两种方法的识别结果均保持不变;人为剔除任意样本中一个共有峰,两种方法的识别共有样本中相应的峰均消失。以上检验表明,按照 1.1 和 1.2 中所述算法建立的 VB 程序无误。

3.3 两种共有峰识别方法比较

分别以时间匹配法和参考光谱法对 8 个荒漠肉苁蓉正品和全部 12 个供试品进行了共有峰的识别,得到共有峰作为指纹图谱。详见表 3。由表可见,时间匹配法得到的共有峰始终比参考光谱法多,说明存在一些峰虽然相对保留值处在可能匹配的范围,但是最大吸收波长差异很大(超过 20 nm),并非同一物质。

本实验中, N_t 最佳取值是 0.15,此时两种方法判定的共有峰均比较多。考虑到仪器方法的固有波动性, N_t 取值过小会导致共有峰数目少,即使同一物质的峰也可能错判为非共有峰; N_t 取值过大则会造成误判,前面部分峰的误判可能造成后面部分继续判定错误,反而减少共有峰总数。另外,8 个正品荒漠肉苁蓉样品的共有峰数目明显高于全部共有样本,说明正品之间比较一致,与其它品种之间存在较大差异。

表 3 两种识别方法指纹图谱共有峰数的比较
Table 3 Comparison of common peak numbers identified by the two methods

N_t	<i>Cistanche. deserticola</i> ^[8]		All samples ^[12]	
	Time matching	Time-spectrumr matching	Time matching	Time-spectrumr matching
0.005	1	1	0	0
0.01	3	2	1	1
0.05	11	7	4	2
0.1	12	10	8	4
0.12	12	11	8	5
0.15	13	11	9	5
0.2	13	10	10	5
0.3	16	9	10	5
0.4	17	6	12	5
0.5	17	7	12	6

传统的人工识别,对于出峰比较密集,并且形状大小差异不大的区域,难以作出准确判断,只能主观判定,用计算机处理则可避免这种主观差异。例如在 6[#] 和 7[#] 样品色谱图中,6[#] 样品的第 24 号峰与 7[#] 样品的第 19 号峰、20 号峰的相对保留值和峰形都比较接近,均有可能匹配,人工识别困难。时间匹配法,判别 6[#] 样品的第 24 号峰与 7[#] 样品的 20 号峰是共有峰,但比较各色谱峰的光谱图,判定 7[#] 样品的 20 号峰是 6[#] 样品第 24 号峰的共有峰,用程序处理则迅速、客观(表 4)。

再如 3[#] 和 8[#] 样品,用时间匹配和参考光谱法分别得到 20 和 19 个共有峰,其中时间匹配法多出的一个峰是将最大吸收波长分别是 230.3 nm 和 328.6 nm 的两个峰判别为共有峰,虽然两峰的相对保留值的差值在 Nt 范围内,时间匹配法认定为共有峰,但这两个峰明显不是同一物质,详见表 4。这进一步表明,同时考虑相对保留值和最大吸收波长的参考光谱法可以更准确的识别共有峰。

表 4 共有峰识别方法比较表
Table 4 Comparison of identification methods

Sample	Peak No.	Relative retention time(min)	Area percent (%)	max (nm)
6 [#]	24	1.244	0.9452	230.3
7 [#]	19	1.227	0.9648	228.0
	20	1.248	0.2111	220.8
3 [#]	30	1.585	0.3319	230.3
8 [#]	26	1.641	0.8174	328.6

4 结论

按照相对保留值最接近建立的时间匹配识别法以及同时参考相对保留值与二极管阵列检测器的光谱信息(最大吸收波长或者整个光谱)建立的参考光谱识别法,只需要人为设置有效峰大小下限、时间相对差值上限和最大吸收波长差值上限三个参数,VB 程序编制的指纹分析系统就能快速对所有样本进行共有峰识别和模糊数学处理。使指纹图谱的建立过程客观化,避免了除实验环节外的人为影响,有利于更好的发挥质量评价的作用,并且节省大量时间。在使用二极管阵列检测器的多维信息检测器时,参考光谱法(或者参考其它多维信息)可以更准确的识别共有峰。

参考文献:

- [1] GAO Yue(高月), LV Qiu-jun(吕秋军), LIU Yong-xue(刘永学), CHEN Peng(陈鹏), TAN Hong-lin(谭洪玲). Chinese New Drugs Journal(中药复方物质基础的研究) [J], 2001, 9(5):307.
- [2] LIANG Yi-zeng(梁逸曾). Traditional Chinese Drug Research & Clinical Pharmacology(中药新药与临床药理) [J], 2001, 12(3):196.
- [3] FENG Xue-song(冯雪松), DONG Hong-ye(董鸿晔). Progress in Pharmaceutical Sciences(药学进展) [J], 2002, 26(4):198.
- [4] LIANG Xin-miao(梁鑫淼), XU Qing(徐青). Chinese Journal of Natural Medicines(中国天然药物), 2003, 1(1):57.
- [5] LI Yong-guo(李永国), WANG Zheng-tao(王峥涛). Traditional Chinese Drug Research & Clinical Pharmacology(中药新药及临床药理) [J], 2006, 16(1):15.
- [6] Lu Peizhang, Liang Xinmiao, Xiao Hongbin. Progress in Chemistry(化学进展) [J], 1999, 11(2):199.
- [7] Editorial Committee of the Pharmacopoeia of People's Republic of China(中华人民共和国药典委员会). Guidance of Traditional Chinese Medicine Chromatographic Fingerprint(Trial Version)(中药注射剂色谱指纹图谱实验研究技术指南(试行)) [S]. 2002.
- [8] Wang L C, Cao Y H, Xing X P. Chromatographia[J], 2005, 62(5/6):283.
- [9] XIA Peng-xiao(夏鹏霄), YANG Han-chun(杨瀚春), YUAN Hong-ze(袁洪泽), GUO Dong-huan(郭东桓). Chinese Journal of Drug Application and Monitoring(中国药物应用与监测) [J], 2006(5):29.
- [10] CHEN Min-jun(陈闽军), CHENG Yi-yu(程翼宇). Chinese J. Anal. Chem.(分析化学) [J], 2003, 31(5):513.
- [11] CHAI Yi-feng(柴逸峰), LUO Guo-an(罗国安), QIAO Shan-lei(乔善磊), ZHU Zhen-yu(朱臻宇), LU Feng(陆峰), WANG Jian-jun(王建军). Chinese J. Anal. Chem.(分析化学) [J], 2005, 33(11):1599.
- [12] ZHU Zhen-yu(朱臻宇), QIAO Shan-lei(乔善磊), ZHANG Hai(张海), LOU Zi-yang(娄子洋), CHAI Yi-feng(柴逸峰). Acad J. Sec. Mil. Med. Univ. (第二军医大学学报) [J], 2007, 28(2):183.
- [13] NI Li-jun(倪力军), WANG Guo-dong(王国东), GUO Jia(郭佳), ZHANG Li-guo(张立国). Chinese J. Anal. Chem.(分析化学) [J], 2006, 34(10):1454.
- [14] WANG Long-xing(王龙星), XIAO Hong-bin(肖红斌), LIANG Xin-miao(梁鑫淼), BI Kai-shun(毕开顺). Acta Pharmaceutica Sinica(药学学报) [J], 2002, 37(9):713.

- [15] Fan X H, Cheng Y Y, Ye Z L, Lin R H, Qian Z Z. *Analytica Chimica Acta* [J], 2006, **555**:217.
- [16] LI Bo-yan (李博岩), LIANG Yi-zeng (梁逸曾), XIE Pei-shan (谢培山), YU Ru-qing (俞汝勤). *Chinese J. Anal. Chem.* (分析化学) [J], 2003, **7**(7):799.
- [17] CHEN Di-zhao (陈迪钊). *Journal of Shangqiu Teachers College* (商丘师范学院学报) [J], 2000, **16**(14):74.
- [18] WANG Ya-li (王亚丽), LIANG Yi-zeng (梁逸曾), HU Yun (胡芸), LI Bo-yan (李博岩), ZENG Zhong-da (曾仲大), XU Shu-ping (徐抒平). *Chemical Journal of Chinese Universities* (高等学校化学学报) [J], 2006, **17**(2):222.
- [19] GONG Fan (龚范). *University Chemistry* (大学化学) [J], 2005, **20**(5):52.
- [20] Jiangsu New Hospital (江苏新医院). *Chinese Traditional Medicine Glossary* (中药大辞典) [M]. Shanghai (上海): Shanghai Science and Technology Press (上海科学技术出版社), 1977, 895.
- [21] GONG Li-dong (龚立冬), CAO Yu-hua (曹玉华), HOU Jian-xia (侯建霞). *Chinese Journal of Chromatography* (色谱) [J], 2007, **25**(2):280.
- [22] XIE Jie-na (谢洁娜), ZHAO Ming-bo (赵明波), WU Feng-wei (吴凤薇), TU Peng-fei (屠鹏飞). *Chinese Traditional and Herbal Drug* (中草药) [J], 2005, **36**(2):268.

Automatic Identification of the Common Peaks in the Chromatographic Fingerprints of Traditional Chinese Medicines

FENG Qi-yu, CAO Yu-hua*, GONG Li-dong

(School of Chemical and Material Engineering, Jiangnan University, Wuxi 214122)

Abstract: Automatic identification of the common peaks in the chromatographic fingerprints of traditional Chinese medicines was established based on two methods of time matching and time-spectrum-matching identification which are based on the relative retention time and spectrum of each peak, respectively. Automatic identification was accomplished by the self-built program utilizing VB programming language. Two methods have been successfully applied in the chromatographic fingerprints of *Cistanche salsa*. It showed that if the parameters including the lowest peak area, the permissible deviation of retention time and upper limit of difference of the maximum absorption wavelength were set, the fingerprint analysis system could rapidly identify the common peaks and do similarity analysis automatically. The fingerprint established with this system was more accurate and consistent, and avoided personal bias. Compared with time-spectrum-matching identification, time matching was more accurate and reliable.

Keywords: Fingerprint analysis; Common peak identification; *Cistanche salsa*; High performance Liquid chromatography-photodiode array detector (HPLC-DAD)