

中药石斛质量的化学模式识别

张亮 马国祥* 张正行 徐国钧* 安登魁

(中国药科大学药物分析研究室; * 生药学教研室, 南京 210009)

摘要 依据任一物质在一定波长下的吸收系数为常数这一基本事实, 以每一波长与另一波长的吸收值比值作指标, 用 Shannon 信息理论对全部样品(32个)计算每一波长通道的信息量, 选取信息量最大的10个通道用作分类特征, 采用主成分分析对中药石斛质量进行了研究, 分类结果与植物形态学鉴定结果基本一致。本法不经分离直接以药材粉末氯仿浸出液的紫外光谱进行分类, 简便、快速, 为从化学成分角度研究中药质量提供了一种新方法。

关键词 主成分分析; 紫外分光光度法; Shannon 信息理论; 石斛; 化学模式识别

中药石斛早在《神农本草经》中就有记载, 并列为上品, 具有滋阴清热、生津溢胃、润肺止咳的功效。商品石斛来源于兰科(Orchadaceae)石斛属、金石斛属、石仙桃属及石豆兰属等多种植物的茎。来源复杂, 现已查明约30个品种在市场上流通。这些来源不同的品种, 某些植物形态差异较大, 某些又具有较大的相似性。因此, 鉴定评价工作往往比较困难。现代科学表明植物化学成分直接影响药材的质量与疗效。因此, 应该从化学成分角度研究中药石斛的质量。

某些植物的溶剂浸出液的紫外光谱(UV)有相对稳定性与特征性, 其光谱吸收带的波长强度对任一物质通常为一常数, 可作为分类和鉴别特征。但是, 紫外光谱主要反映分子中羰基和基团性质, 它的谱形较简单, 吸收峰宽且成带状, 对具有相似光谱的物质就难以区别。有人用每间隔5 nm 取一吸收值作为指标进行中药质量评价⁽²⁾, 但该法间隔太大可能丢失部分谱变化信息, 而且吸收值随浓度而变化, 为不稳定性特征指标, 数量标度缺少可比性。本文用1 nm 间隔测吸收值, 每波长与特定波长的吸收值比值作为指标, 采用 Shannon 信息理论, 对样品计算每一波长通道的信息量, 取富含信息量的通道作为分类指标, 最后, 运用主成分分析法(PCA)进行综合分析, 获得满意结果。

实 验 部 分

仪器与试剂

UV-260型紫外—可见分光光度计(岛津)。石斛原药材部分系作者采集并鉴定, 部分为商品, 详见表1。化学试剂均为分析纯。

Tab 1 Objects used in principal component analysis (PCA)

No	Name	Collection place
1	<i>Dendrobium linavianum</i>	Yunnan, Xishuang Banna
2	<i>D. aphyllum</i>	Ditto
3	<i>D. crepidatum</i>	Ditto
4	<i>D. hercoglossum</i>	Guangxi, Longliang
5	<i>D. gratiotissimum</i>	Yunnan, Xishuang Banna
6	<i>Ephemeranta lonchophylla</i>	Ditto
7	<i>D. exile</i>	Ditto
8	<i>D. fimbriatum</i>	Ditto
9	<i>D. capillipes</i>	Ditto
10	<i>D. lindleyi</i> var. <i>jenkensis</i>	Hainan
11	<i>D. nobile</i>	Guangxi
12	<i>D. falconeri</i>	Yunnan, Xishuang Banna
13	<i>D. hancockii</i>	Guangxi, Longliang
14	<i>D. yongfuense</i>	Guangxi, Yongfu
15	<i>D. chrysoxum</i>	Yunnan, Xishuang Banna
16	<i>D. moniliforme</i>	Guangxi, Guiliang
17	<i>D. chreyseum</i>	Yunnan, Xishuang Banna
18	<i>D. chrysanthum</i>	Ditto
19	<i>D. chrysanthum</i>	Guangxi, Longliang
20	<i>D. chreyseum</i> var. <i>bulangense</i>	Yunnan, Xishuang Banna
21	<i>D. primulinum</i>	Ditto
22	<i>D. loddigesii</i>	Guangxi, Longliang
23	<i>D. moschatum</i>	Yunnan, Xishuang Banna
24	<i>D. officinale</i>	Guangxi, Guiliang
25	<i>D. bellatulum</i>	
26	<i>Pholidota yunnanensis</i>	Yunnan, Xishuang Banna
27	<i>D. devonianum</i>	Ditto
28	<i>D. aduncum</i>	Guizhou, Xiyi
29	<i>D. densiflorum</i>	Yunnan, Xishuang Banna
30	<i>D. wardianum</i>	Ditto
31	<i>D. jenkinsii</i>	Guangxi
32	<i>D. pendulum</i>	Yunnan, Xishuang Banna

数据采集

取药材粉10 g(原药材量不多时,可酌情减量),氯仿冷浸提取,浸出液经减压浓缩后,转移至10 ml量瓶中,用氯仿稀释至刻度,吸取浸出液3 μ l,转移至锥形瓶中,挥去氯仿,加甲醇溶解残渣(浑浊时需过滤),在UV-260型紫外—可见分光光度计上,于200~360 nm区间,每间隔1 nm测定一个吸收值,得样本原始数据。图1为石斛氯仿浸出液典型的紫外光谱图。

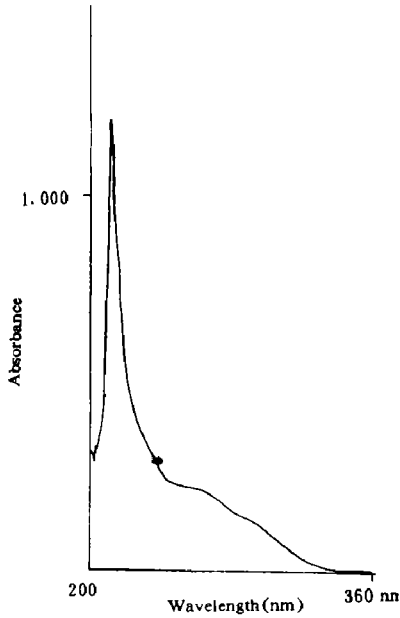


Fig 1 UV spectrum of *Dendrobium* (solvent: methanol).

结果和讨论

样本原始数据是一定浓度样品在测定波长下的吸收值,它随样品浓度大小而变化,不同样本间缺少可比性。因此,需要进行数据转换。假设药材浸出物在一定溶剂中服从比尔定律,那么,对每一浸出物样本,两个测定波长下的吸收值比值为—常数。本文将每一波长下的吸收值与特定波长下的吸收值比值作为分类指标。

特征选择

经处理,每一样本均可得150个特征变量(整数)。对于微型计算机而言,处理样本的高维UV数据比较困难。对于主成分分析(PCA)法,需满足样本数大于变量数3倍以上这一条件。因此,首先须对样本进行降维处理,但是维数(特征数)及其对谱图的分辨率是相互矛盾的两个方面。本文试用 Shannon 信息理论进行特征选取。

Shannon 信息理论是研究从数量上定量描述信息的方法⁽³⁾。信息量的大小主要取决于事件发生的机率和事件间的相互关系,可用下式表示

$$I(j) = - \sum_{i=1}^{m_j} P_j(i) \text{LOG}_2 P_j(i)$$

式中 m_j 为第 j 波长下对所有样本的不连续的最大特征变量值, $P_j(i)$ 为第 j 波长下特征变量值为 i 的样本对所有数据集样本发生的机率, $I(j)$ 为第 j 波长通道对所有样本的信息量之和。各波长通道的信息量计算结果见表2,本文选择 I 值最大的10个波长通道的吸收值作为分类特征,见表3。

Tab 2 Information content of each wavelength channel for all objects

Wavelength (nm)	Bit	Wavelength (nm)	Bit	Wavelength (nm)	Bit	Wavelength (nm)	Bit	Wavelength (nm)	Bit	Wavelength (nm)	Bit
211	2.52	236	3.71	261	3.09	286	3.54	311	3.54	336	2.93
212	2.99	237	3.69	262	3.08	287	3.46	312	3.48	337	3.00
213	3.20	238	3.50	263	3.07	288	3.32	313	3.49	338	2.84
214	3.46	239	3.75	264	3.28	289	3.52	314	3.46	339	2.79
215	3.55	240	3.54	265	3.35	290	3.39	315	3.40	340	2.75
216	3.78	241	3.66	266	3.31	291	3.55	316	3.28	341	2.71
217	3.72	242	3.69	267	3.37	292	3.46	317	3.36	342	2.61
218	3.93	243	3.58	268	3.35	293	3.46	318	3.38	343	2.73
219	3.89	244	3.58	269	3.38	294	3.46	319	3.35	344	2.59
220	3.84	245	3.50	270	3.39	295	3.48	320	3.40	345	2.57
221	3.82	246	3.45	271	3.35	296	3.48	321	3.38	346	2.58
222	3.80	247	3.24	272	3.43	297	3.46	322	3.35	347	2.63
223	3.61	248	3.33	273	3.46	298	3.55	323	3.36	348	2.63
224	3.78	249	3.38	274	3.42	299	3.58	324	3.28	349	2.48
225	3.79	250	3.39	275	3.44	300	3.59	325	3.28	350	2.45
226	3.82	251	3.22	276	3.44	301	3.40	326	3.18	351	2.45
227	3.82	252	3.19	277	3.50	302	3.43	327	3.10	352	2.50
228	3.77	253	3.31	278	3.48	303	3.52	328	3.15	353	2.54
229	3.82	254	3.17	279	3.51	304	3.43	329	3.02	354	2.56
230	3.69	255	3.09	280	3.51	305	3.54	330	3.05	355	2.56
231	3.78	256	3.17	281	3.51	306	3.54	331	2.99	356	2.49
232	3.89	257	3.27	282	3.57	307	3.49	332	2.96	357	2.51
233	3.75	258	3.17	283	3.57	308	3.48	333	2.88	358	2.36
234	3.71	259	3.15	284	3.53	309	3.50	334	2.94	359	2.40
235	3.84	260	3.11	285	3.51	310	3.53	335	2.92	360	2.37

Tab 3 Wavelength channels containing the highest information content

Wavelength (nm)	Bit	Wavelength (nm)	Bit
218	3.93	226	3.82
219	3.89	227	3.82
220	3.84	229	3.82
221	3.82	232	3.89
222	3.80	235	3.84

主成分分析结果

主成分分析是通过适当的数学变换,最大限度地保留原样本集所含原始信息,使新变量成为原变量的线性组合,并寻求主成分来研究样本的一种方法⁽⁴⁾。用于进行多指标综合评价的基本步骤包括:(1)原始指标数据的标准化;(2)求指标数据间的相关矩阵R;(3)求R阵的特征根、特征向量和贡献率等;(4)确定主成分数;(5)将所有样本映射到主因子空间(通常选2或3个主因子)。本文选用 PARVUS 软件包中部分程序结合自编程序,并在386微机上运行。结果如下:第一主成分的特征值 $\lambda_1=9.77$,方差贡献为97.67%,第二主成分的特征值 $\lambda_2=0.17$,方差贡献为1.70%。二个主成分累计方差贡献为99.37%。图2为样本在两个主因子空间的图形分布。

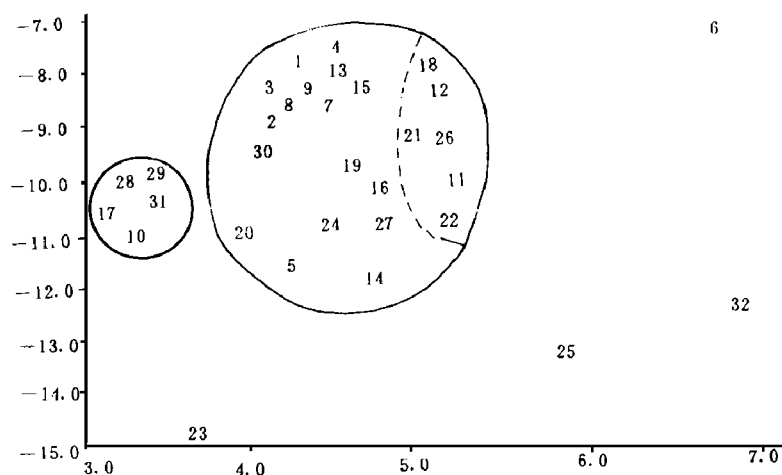


Fig 2 Principal component analysis plot of all objects.

6号为金石斛属(*Ephemerantho*)截叶金石斛(*E. Lonchophylla*),其茎中主要含酚性化合物,有别于其它石斛属植物(主要含生物碱)。

23号为石斛属囊唇石斛(*D. moschatum*),为我国云南南部新分布。植株高大,茎略带紫色,花唇瓣兜囊状,密被绒毛。

25号为石斛属矮石斛(*D. bellatum*),茎矮小,草质被黑毛,迥异于石斛属其它植物。

32号为石斛属肿节石斛(*D. pendulum*),植株矮小,茎肉质,节显著膨大,与本属它种植物形态差异较大。

10,17,28,29,31号样本中,除28号外,茎木质,其中样10,29,31茎呈菱形。该组生物碱含量一般偏低。

以上样本与商品流通中的主流样本差别较大,从PCA图可见:它们与主流样本分离明显,其次,样10,17,28,29,31生物碱含量偏低,样11,12,18,21,22,26生物碱含量偏高,分别聚为一组。由此说明:样6,23,25,32的化学成分与其它品种相比有较大差异,能否继续作为石斛药用,需要进一步研究。

值得一提的是:样26即云南石仙桃为石仙桃属植物,与石斛为不同属,商品流通中却经常出现,PCA分析表明其与石斛属主流样本具有相似的化学成分。因此,作为石斛使用具有一定合理性。

以药材浸出液的紫外光谱为基础进行质量分析,用量少,仪器普及,便于数据采集,与它法比较省时、灵敏、具有较强的实用性。采用主成分分析把紫外光谱作为一整体在以Shannon方程提取特征后进行综合评价,避免了单纯以某一指标或某几个指标(如峰位,峰强度等)进行定性比较的不确定性,是从化学成分角度研究中药质量的又一种具有前途的方法。但是,紫外光谱仅仅反映了部分具有紫外吸收化学成分的行为,全面评价还需结合红外光谱,色谱/质谱等多种技术。作者将另文报告该方面的工作。

参 考 文 献

- 1 江苏新医学院. 中药大辞典. 上海科学技术出版社, 1985: 585~589.
- 2 朱景生, 等. 化学模式识别技术评价中药厚朴质量的研究. 分析测试通报 1988; 7(5): 20.
- 3 Scott DR, et al. Determination of chemical classes from mass spectra of toxic organic compounds by SIMCA pattern recognition and information theory. *Anal Chem* 1986; 58: 881.
- 4 Rose-Pehrsson SL, et al. Detection of hazardous vapors including mixtures using pattern recognition analysis of responses from surface acoustic wave devices, *Ibid* 1988; 60: 2301.

STUDY ON *DENDROBIUM* QUALITY BY CHEMICAL PATTERN RECOGNITION

L Zhang, GX Ma*, ZX Zhang, GJ Xu* and DK An

(*Department of Pharmaceutical Analysis, * Department of Pharmacognosy,
China Pharmaceutical University, Nanjing 210009*)

ABSTRACT Based upon the fact that absorptivity of a substance at a certain wavelength is a constant value, the absorbance ratio at each wavelength to a specified wavelength was used as classification features. The Shannon information content for each wavelength channel was calculated for UV spectra of *Dendrobium* extracts, the 10 wavelength channels with the highest information content were retained as a compressed basis set for principal component analysis (PCA). The results show that PCA is more intrinsic, convenient and feasible than microscopic method.

Key words Principal component analysis; UV spectrometry; Shannon information theory; *Dendrobium*; Chemical pattern recognition