

中药指纹图谱相似度计算的规范化研究

乔善磊¹, 朱臻宇¹, 王建军¹, 柴逸峰^{1*}, 刘艳娥²

(1. 第二军医大学药学院药物分析学教研室, 上海 200433; 2. 解放军总后勤部卫生部药品仪器检验所, 北京 100071)

[摘要] 目的: 探讨影响中药色谱指纹图谱相似度计算的保留时间校正和内标峰选择。方法: 对金银花药材指纹图谱进行保留时间和峰面积校正, 分别使用相关系数法和夹角余弦法计算处理前后图谱间相似度。结果: 保留时间校正前后相似度结果差异较大, 内标峰选择与否和选择不同内标峰的各图谱与对照图谱的相似度均有差异。结论: 相似度计算中, 内标峰选择和保留时间校正须规范化, 才能获得稳定的计算结果。

[关键词] 中药指纹图谱; 相似度; 计算

[中图分类号] R 282 **[文献标识码]** A **[文章编号]** 0258-879X(2004)10-1114-03

Standardization study on similarity calculation for traditional Chinese medicines fingerprints

QIAO Shan-Lei¹, ZHU Zhen-Yu¹, WANG Jian-Jun¹, CHAI Yi-Feng^{1*}, LIU Yan-E² (1. Department of Pharmaceutical Analysis, School of Pharmacy, Second Military Medical University, Shanghai 200433, China; 2. Institute for Drug and Instrument Control, Health Department, General Logistics Department of PLA, Beijing 100071)

[ABSTRACT] **Objective:** To discuss the selection of retention time and internal standard peak in similarity calculation for traditional Chinese medicines (TCM) fingerprints. **Methods:** The similarities of honeysuckle chromatography fingerprints were calculated by the methods of correlation coefficients and angle cosine with or without retention time correction and signal standardization. **Results:** When similarity was calculated between control and sample chromatography fingerprints, differences were found using different data treatments procedure, for example, correlating retention time and selecting internal standard peak. **Conclusion:** In order to get a stable result in similarity calculation for TCM fingerprints, the retention time and internal standard peak must be standardized in treatment procedure.

[KEY WORDS] traditional Chinese medicines fingerprint; similarity; calculation

[Acad J Sec Mil Med Univ, 2004, 25(10): 1114-1116]

* 中药指纹图谱技术被认为是中药质量控制的一种有效手段^[1]。整体性和模糊性是中药指纹图谱的基本属性^[2]。模糊性强调的是对照样品与待测样品指纹图谱的相似性, 而不是完全相同; 整体性是强调完整地比较色谱的特征“面貌”, 而不是将其“肢解”。使用相似度来描述指纹图谱之间的相似情况, 将指纹图谱作为一个整体来考虑, 而不是局限于某一个物质的含量的大小, 与当前使用的中药质量控制方法相比有一定优越性。夹角余弦法和相关系数法是中药指纹图谱相似度计算的主要方法, 这 2 种方法都可以量化描述指纹图谱间的相似情况。相似度计算的运算量较大, 需借助计算机辅助进行。国家食品药品监督管理局在 2002 年推出了两个中药指纹图谱相似度计算软件, 以方便研究者计算指纹图谱之间的相似度。

本课题在相关研究基础上, 以开发的《中药指纹图谱工作站》软件, 作为中药指纹图谱实验数据处理的工具。

1 原理和方法

中药指纹图谱的相似度计算是将一张指纹图谱作为一个 n 维向量 $X(x_1, x_2, x_3, \dots, x_n)$, 这个向量可以是指纹图谱的原始数据点, 也可以是指纹图谱的色谱峰积分信息。本研究相似度计算使用原始数据点进行。两张指纹图谱之间的相似度是对两指纹图谱向量 A 、 B 使用一定的公式计算所得。夹角余弦法和相关系数法的计算公式参见文献^[3]。

指纹图谱实验中, 可能发生保留时间的漂移, 对相似度的计算产生影响, 因此需要对保留时间进行校正。本研究所使用的保留时间校正方法为两点法, 即每个图谱选择两个保留时间参照峰进行校正, 其

* [基金项目] 上海市科委发展基金课题(01DJ19012)。

[作者简介] 乔善磊(1974-), 男(汉族), 硕士

E-mail: aqiao1974@sina.com

* Corresponding author. E-mail: yfchai@smmu.edu.cn

校正公式为: $t_{校正后} = \bar{t}_{前} + (t_{未校正} - \bar{t}_{前}) \frac{\bar{t}_{后} - \bar{t}_{前}}{t_{后} - t_{前}}$ 。两点法校正的实质是将待校正色谱图中保留时间参照峰对齐后进行伸缩变换。

根据 k 个指纹图谱向量 X_1, X_2, \dots, X_k , 求得的中位数或平均数向量称为对照图谱。对照图谱是评判指纹图谱是否合格的一个依据^[4]。

信号强度的校正是使用内标峰对指纹图谱进行标准化处理, 对于一个指纹图谱向量 X_i , 其内标峰面积为 A_i , 其校正公式为: $X_{校正} = X_i / A_i$ 。

本研究所采用的数据为金银花药材高效液相指纹图谱, 由解放军总后勤部卫生部药品仪器检验所提供^[5]。图谱间的相似度计算使用第二军医大学和清华大学联合开发的《中药指纹图谱工作站》软件。

2 结果和讨论

2.1 对图谱直接计算相似度 从金银花高效液相指纹原始图谱(图 1)中可以直观看出 3、4、5 号样品与 1、2 号有一定的差别, 但通过夹角余弦与相关系

数法所反映的相似度趋势基本相同(表 1); 从相似度计算结果来看, 样品 1、2 属于离群样本, 使用平均数计算的对照图谱中包含较多离群样本(样品 1、2)信息, 中位数的对照图谱则包含较少的离群样本信息, 因此两种对照图谱与离群样本的相似度有较大差别。

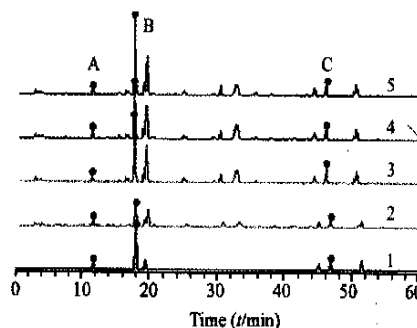


图 1 金银花高效液相色谱指纹图谱
Fig 1 The HPLC fingerprints of honeysuckle 1-5: Chromatograms of honeysuckle from different original producing areas; A, B, C: Peaks used in retention time correction and signal standardization

表 1 金银花高效液相指纹原始图谱相似度

Tab 1 Similarities of HPLC fingerprints of honeysuckle

Sample No.	Sample No.				Reference fingerprint	
	1	2	3	4	Mean	Median
1	-	-	-	-	0.41(0.39)	0.18(0.16)
2	0.76(0.76)	-	-	-	0.42(0.39)	0.21(0.17)
3	0.15(0.13)	0.18(0.14)	-	-	0.95(0.95)	0.99(0.99)
4	0.13(0.11)	0.16(0.13)	0.98(0.98)	-	0.94(0.94)	0.98(0.98)
5	0.17(0.15)	0.21(0.17)	0.99(0.99)	0.96(0.95)	0.96(0.95)	0.98(0.98)

Data in and out bracket are the similarities calculated by cosine formula and correlation coefficient formula, respectively

2.2 对图谱保留时间进行校正后计算相似度 表 2 结果在相似度计算前使用 B、C 两组峰对图谱保留时间进行校正, 但没有校正信号强度。保留时间校正前后的相似度差异较大, 说明保留时间校正为相似度计算必要的步骤, 同时, 保留时间校正也为以后建立规范的样本矩阵进行模式识别打下基础; 研究表

明, 样品 1、2 的离群表现, 在很大程度上是由于没有进行保留时间校正造成的。在不存在离群样本的情况下, 中位数和平均数对照图谱与样品的相似度情况基本相同。国家食品药品监督管理局要求的一些统计量如 RSD^[4], 是在使用平均数时用到的统计量, 因此建议只采用平均数向量作为对照图谱。

表 2 保留时间校正后金银花指纹图谱的相似度

Tab 2 Similarities after retention time correction of HPLC fingerprints of honeysuckle

Sample No.	Sample No.				Reference fingerprint	
	1	2	3	4	Mean	Median
1	-	-	-	-	0.87(0.87)	0.83(0.83)
2	0.75(0.75)	-	-	-	0.94(0.94)	0.94(0.94)
3	0.83(0.83)	0.93(0.93)	-	-	0.99(0.99)	0.99(0.99)
4	0.81(0.81)	0.93(0.92)	0.99(0.99)	-	0.99(0.99)	0.99(0.99)
5	0.83(0.83)	0.95(0.94)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)

Data in and out bracket are the similarities calculated by cosine formula and correlation coefficient formula, respectively

2.3 对保留时间和信号强度进行校正后计算相似度。表3、表4结果在相似度计算前使用B、C两组峰对图谱保留时间进行校正,并分别使用A峰、B峰进行信号强度校正。使用内标峰校正图谱,并不改变样品间的相似度(由夹角余弦和相关系数的计算公式也可以看出这一点),夹角余弦和相关系数相似度对样品的浓度变化是不敏感的,只对样品的组成物之间的相对比例变化敏感。选择了内标峰后,所形成的对照图谱与不选内标峰时不同,对照图谱与样

品的相似度也不相同;选择不同的内标峰,形成的对照图谱不同,因此与样品的相似度也不相同。在色谱方法中,使用内标峰进行校正,以建立规范的对照图谱向量,并且在实验设计时就应当考虑内标物质的加入问题,不应当在数据处理时才选择内标峰,从而提高图谱间的可比性。如果试验目的不是建立对照图谱指纹图谱,而仅仅是计算各样品之间的相似度或样品与已经建立的对照图谱的相似度,没有必要使用内标峰。

表3 保留时间校正并使用A峰校正信号强度后相似度

Tab 3 Similarities after retention time correction and signal standardized by peak A

Sample No.	Sample No.				Reference fingerprint	
	1	2	3	4	Mean	Median
1	-	-	-	-	0.86(0.86)	0.83(0.83)
2	0.75(0.75)	-	-	-	0.94(0.94)	0.95(0.95)
3	0.83(0.83)	0.93(0.93)	-	-	0.99(0.99)	0.99(0.99)
4	0.81(0.81)	0.93(0.92)	0.99(0.99)	-	0.99(0.99)	0.99(0.99)
5	0.83(0.83)	0.95(0.94)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)

Data in and out bracket are the similarities calculated by cosine formula and correlation coefficient formula, respectively

表4 保留时间校正并使用B峰校正信号强度后相似度

Tab 4 Similarities after retention time correction and signal standardized by peak B

Sample No.	Sample No.				Reference fingerprint	
	1	2	3	4	Mean	Median
1	-	-	-	-	0.87(0.87)	0.84(0.84)
2	0.75(0.75)	-	-	-	0.95(0.95)	0.94(0.94)
3	0.83(0.83)	0.93(0.93)	-	-	0.99(0.99)	0.99(0.99)
4	0.81(0.81)	0.93(0.92)	0.99(0.99)	-	0.98(0.98)	0.99(0.99)
5	0.83(0.83)	0.95(0.94)	0.99(0.99)	0.99(0.99)	0.99(0.99)	0.99(0.99)

Data in and out bracket are the similarities calculated by cosine formula and correlation coefficient formula, respectively

本研究通过对中药指纹图谱进行不同处理,得到的相似度计算结果不同。中药指纹图谱相似度作为中药质量控制的手段,必须对计算方法和步骤加以规范化,选择合适的内标峰,并以此内标峰为基准,采用适当的方法校正保留时间和信号强度,以获得稳定的计算结果。

[参考文献]

- [1] 任德权. 中药指纹图谱质控技术的意义与作用[J]. 中药材, 2001, 24(4): 235-237.
 [2] 刘树深, 易忠胜. 基础化学计量学[M]. 北京: 科学出版社,

1999: 131.

- [3] 谢培山. 色谱指纹图谱分析是中草药质量控制的可行策略[J]. 中药新药与临床药理, 2001, 12(3): 141-151.
 [4] 国家药品监督管理局. 中药注射剂指纹图谱研究的技术要求(暂行). 国药管注[2000]348号.
 [5] 刘艳娥, 吴玖涵, 吴玉田, 等. 金银花药材高效液相指纹图谱研究[J]. 解放军药学学报, 2003, 19(3): 161-164.
 Liu YE, Wu MH, Wu YT, et al. Study on fingerprint of honeysuckle by high performance liquid chromatography [J]. Jiefangjun Yaoxue Xuebao, 2003, 19(3): 161-164.
 [收稿日期] 2004-02-01 [修回日期] 2004-06-23
 [本文编辑] 尹茶