

# 蚌兰花挥发油化学成分色谱保留值的构效关系研究

冯长君, 石春玲, 李鸣建  
(徐州工程学院化学化工学院, 江苏 徐州 221008)

**摘 要:** 基于著名的 Kier 连接性指数( $^mX^V$ )及电性距离矢量( $M_d$ ), 使用最佳变量集回归(LBR)建立蚌兰花中 20 种有机成分的气相色谱保留时间( $t_R$ )的四元数学模型, 其传统相关系数( $R^2$ )为 0.963, 逐一剔除法(LOO)交叉验证系数( $Q$ )为 0.832。该模型具有高度稳健性与良好的预测能力, 影响有机物气相色谱保留时间的主要因素是分子大小与空间形状。结果表明, Kier 逆指数对有机物气相色谱保留时间的表征是合理的、有效的, 所建模型能较好解释其递变规律。

**关键词:** 蚌兰花; 保留时间; 连接性指数; 电性距离矢量; 定量结构-保留相关(QSRR)

## Quantitative Structure-retention Relationship Studies of Chemical Constituents in Volatile Oil from *Rhoeo spathacea* (Sw.) Stearn Flowers

FENG Chang-jun, SHI Chun-ling, LI Ming-jian  
(School of Chemistry and Chemical Engineering, Xuzhou Institute of Technology, Xuzhou 221008, China)

**Abstract:** On the basis of Kier's molecular connectivity index ( $^mX^V$ ) and electronegativity distance vector ( $M_d$ ), the retention time of 20 organic components in volatile oil from *Rhoeo spathacea* (Sw.) Stearn flowers was determined and a four-variable model (QSRR) of gas chromatographic retention time ( $t_R$ /min) for 20 chemical components was established from  $^mX^V$ ,  $M_d$  by leaps-and-bounds regression (LBR). Traditional correlation coefficient ( $R$ ) and cross-validation correlation coefficient ( $Q$ ) of leave-one-out (LOO) were 0.963 and 0.832, respectively. The model was highly reliable and had favorable prediction capability. The dominant factors for retention time were molecular size and spatial shape of organic molecules. The Kier's converse indices and electronegativity distance vector exhibited a good rationality and efficiency in characterizing retention time of organic compounds. This model could elucidate the change trend of retention indices for organic components.

**Key words:** *Rhoeo spathacea* (Sw.) Stearn; gas chromatographic retention time; connectivity index; electronegativity distance vector; quantitative structure-retention relationship (QSRR)

中图分类号: R284.1

文献标识码: A

文章编号: 1002-6630(2010)19-0042-03

蚌兰花为多年生的肉质草本植物, 属鸭跖草科。其花味甘淡、性凉、入肺、脾经。有清热润肺、止咳化痰、凉血止痢的功效, 主治内伤肺癆、淋巴结核、支气管炎、肺热燥咳以及吐血、衄血、血痢、便血等症, 故有“血见愁”俗名。黄丽莎等<sup>[1]</sup>采用水蒸气蒸馏法提取蚌兰花鲜花的挥发油, 并通过 GC-MS 联用技术对其化学成分进行分析研究, 鉴定出了相对含量大

于 0.3% 的 20 种化学成分(表 1), 主要成分为天竺葵醛、棕榈酸、壬酸、月桂酸、异香兰醛、葡萄花酸、羊脂酸、亚油酸等。现代医学研究表明, 不饱和脂肪酸(如亚油酸等)不但具有营养脑细胞、调节交感神经、显著的降血脂作用及抗炎效果, 而且对于降低血液中胆固醇的含量、预防高血压和动脉粥样硬化具有明显效果<sup>[2]</sup>。因此, 长期食用, 利于人体生理平衡调节, 益于身体健康。

收稿日期: 2010-01-13

基金项目: 国家自然科学基金项目(20776149); 环境模拟与污染控制国家重点联合实验室基金项目(KJ2007001); 徐州工程学院培育课题(XKY2008313)

作者简介: 冯长君(1954—), 男, 教授, 本科, 主要从事有机物构效关系研究。E-mail: fengcj@xzit.edu.cn

表1 蚌兰花挥发油中20种有机组分的拓扑指数及其保留时间  
Table 1 Topological indices and retention time ( $t_R$ ) of 20 organic components in volatile oil from *Rhoeo spathacea* (Sw.) Stearn flowers

编号	化学成分	$^0X_p^v$	$M_{14}$	$M_{22}$	$M_{82}$	$t_R/\text{min}$		
						Exp	Cal	Err
1	葡萄糖酸	5.8910	5.5432	0.0000	0.0000	7.93	7.24	0.69
2	天竺葵醛	6.9353	12.3160	0.0000	0.0000	8.03	10.62	-2.59
3	羊脂酸	6.5981	8.1950	0.0000	0.0000	9.43	9.07	0.36
4	壬酸	7.3052	11.0630	0.0000	0.0000	11.34	10.95	0.39
5	黄樟脑	6.7472	11.5340	8.4748	1.3372	11.37	11.82	-0.45
6	十一醛	8.3496	18.6030	0.0000	0.0000	11.50	14.52	-3.02
7	羊蜡酸	8.0123	14.0820	0.0000	0.0000	12.91	12.87	0.04
8	4,5-二甲基-2-十五基-1,3-二氧戊烷	15.4480	39.0970	5.2657	2.4131	13.45	13.20	0.25
9	异香兰醛	6.0731	8.6464	1.9481	0.0000	13.72	12.43	1.29
10	香叶基丙酮	9.8914	6.8209	0.0000	0.0000	14.30	14.35	-0.05
11	肉桂酸	5.8969	30.9930	0.0000	0.0000	14.64	13.28	1.36
12	十一酸	8.7194	17.2130	0.0000	0.0000	14.91	14.82	0.09
13	月桂酸	9.4265	20.4280	0.0000	0.0000	17.09	16.79	0.30
14	雪松醇	11.2150	7.6257	0.0000	0.0000	17.97	16.80	1.17
15	十三烷酸	10.1340	23.7070	0.0000	0.0000	19.02	18.77	0.25
16	肉豆蔻酸	10.8410	27.0380	0.0000	0.0000	21.07	20.76	0.31
17	植酮	14.0040	14.5760	0.0000	0.0000	22.39	23.20	-0.81
18	棕榈酸	12.2550	33.8150	0.0000	0.0000	25.18	24.78	0.40
19	亚油酸	13.1500	48.9440	0.0000	0.0000	28.14	29.89	-1.75
20	硬脂酸	13.6690	40.7060	0.0000	0.0000	30.59	28.82	1.77

注: Exp 为实验值; Cal 为计算值; Err 为计算误差。

Randic<sup>[3]</sup>最早将分子连接性指数( $^mX$ )用于醇类气相色谱保留指数(RI)的研究,随后许多学者相继建立了烃及其衍生物的定量结构-色谱保留指数相关性模型(QSRR)<sup>[4-8]</sup>,现已成为色谱科学领域中十分活跃的课题。本实验依据Kier等<sup>[9]</sup>的分子连接性指数( $^mX^v$ )及Liu等<sup>[10-12]</sup>的电性距离矢量(MEDV,以 $M_d$ 表示),20种蚌兰花挥发油化学成分的气相色谱保留时间( $t_R$ )<sup>[1]</sup>拟合,经最佳变量子集回归(Leaps-and-Bounds regression, LBR)得四元QSRR模型,其复相关系数( $R^2$ )为0.963,逐一剔除法(leave-one-out, LOO)交互校验复相关系数( $Q^2$ )为0.832。结果表明所建模型具有良好的预测能力与总体稳健性。

## 1 材料与方法

### 1.1 材料

蚌兰花挥发油中20种化学成分的保留时间( $t_R/\text{min}$ )取自文献[1]。

### 1.2 拓扑指数构建方法

常用的QSAR方法有辛醇-水分配系数法、线性溶剂化能法及拓扑指数法等,但以后者最为简便,因它几乎不需查找任何化学参数。拓扑指数是以分子的二维结构为基础,通过图论的方法构建结构参数,以反映分子中原子间的连接方式与次序。迄今已报道了400余种拓扑指数,其中以Kier等<sup>[9]</sup>的价连接性指数( $^mX^v$ )的应

用最广泛。Kier等<sup>[9]</sup>在分子隐氢图的邻接矩阵基础上,定义分子的价连接性指数( $^mX^v$ )。

$$^mX^v = \sum (\delta_i^v \delta_j^v \dots)^{1/2} \quad (1)$$

式中: $^mX^v$ 是一个指数体系,由 $(m+1)$ 个指数组成; $m$ 表示相应指数的阶数, $m=0,1,2,\dots$ ;  $t$ 代表子图的类型,常用4种子图链、星、星-链、环状子图,依次对应 $t$ 为p、c、pc及ch。 $\delta_i^v$ 为原子点价,表征原子结构信息,定义式为:

$$\delta_i^v = \frac{m_i - h_i}{M_i - m_i - 1} \quad (2)$$

式中: $M_i$ 、 $m_i$ 为非氢原子 $i$ 的电子总数及价电子数, $h_i$ 为与非氢原子 $i$ 直接键合的氢原子数。本研究利用MATLAB软件计算了11种 $^mX^v$ : $^0X_p^v$ 、 $^1X_p^v$ 、 $^2X_p^v$ 、 $^3X_p^v$ 、 $^4X_p^v$ 、 $^5X_p^v$ 、 $^3X_c^v$ 、 $^4X_c^v$ 、 $^5X_{pc}^v$ 、 $^5X_{pc}^v$ 、 $X_{ch}^v$ 。

Liu等<sup>[10-12]</sup>考察了多种著名拓扑指数的局限,提出能够较为全面反映分子的拓扑、几何及电性特征的电性距离矢量。电性距离矢量的具体计算方法见文献[10-12],例如,1,2-二氯乙烷的 $M_d$ 计算过程如下:

首先计算此分子中二种非氢原子的原子属性( $I_i$ ,  $i$ 代表非氢原子类型):—C—(第二类原子),—Cl(第十三类原子)的 $I$ 值依次为1.5000、1.9108。其次是计算相应的电性状态指数( $E_i$ )为:0.9865、2.4243。第三步是计算电性距离矢量( $M_d$ ,  $d$ 为电性距离矢量的序数):这两种非氢原子两两组合共有3种 $M_d$ : $M_{14}$ (即 $M_{2-2}$ )、 $M_{25}$ (即 $M_{2-13}$ )及 $M_{91}$ (即 $M_{13-13}$ ),其值依次为0.9732、5.9789、0.6530。

本研究化合物分子中共有6种原子类型: $n=1,2,3,4,9,10$ ,它们两两组合构成21个 $M_d$ 指数。利用程序计算了蚌兰花20种化学成分的 $M_d$ 指数。

### 1.3 多元统计回归方法

将每个挥发性化学成分分子的33种结构描述符(12种连接性指数与21种电性距离矢量)作为自变量,相应的 $t_R$ 为因变量输入Mintab14。首先采用其中最佳子集回归(Leaps-and-Bounds regression)程序选择最佳变量组合,建立相应定量结构-保留指数相关性(QSRR)模型,然后应用逐一剔除法对模型的预测能力及稳健度进行检验,以交叉验证相关系数( $Q^2$ )予以评价。

## 2 结果与分析

### 2.1 蚌兰花挥发性成分的QSRR模型

20个蚌兰花挥发性成分的气相色谱保留时间( $t_R$ )数据来自文献[1],将其和11种分子价连接性指数( $^mX^v$ )、21种电性距离矢量( $M_d$ )一起输入Minitab软件系统,经最佳变量子集回归建立的QSRR模型见表2。其中 $R^2$ 、 $R_{adj}^2$ 、 $Q^2$ 、 $S$ 、 $F$ 分别为相关系数、校正判定系数(以消除自

变量个数及样本容量对判定系数的影响)、逐一剔除法的交叉验证系数、估计标准误差、Fischer 检验值。

表 2  $t_R$  与  $^mX_p^v$ ,  $M_d$  的最佳变量子集回归结果

Table 2 Leaps-and-bounds regression for  $t_R$  as a function of  $^mX_p^v$ , and  $M_d$

试验号	$R^2$	$R_{adj}^2$	$Q^2$	$S$	$F$	变量
1	0.551	0.526	0.320	4.430	22.073	$^0X_p^v$
2	0.737	0.706	0.000	3.490	23.789	$^0X_p^v$ , $M_{82}$
3	0.850	0.822	0.453	2.719	30.133	$^1X_p^v$ , $M_{82}$ , $M_{22}$
4	0.963	0.953	0.832	1.391	97.897	$^0X_p^v$ , $M_{82}$ , $M_{22}$ , $M_{14}$
5	0.980	0.973	0.794	1.122	122.039	$^1X_p^v$ , $M_{82}$ , $M_{33}$ , $M_{77}$ , $X_{ch}^v$

由表 2 可见,随着模型中变量数增多,其  $Q^2$  呈锯齿状变化,先是增大至 0.832,而后下降,说明该四元数学模型具有最好的稳定性及预测能力,相应模型参数为:

$$t_R = -4.113 + 1.703^0X_p^v + 0.237M_{14} + 2.219M_{22} - 12.218 \times M_{82} \quad (3)$$

$$n=20, R^2=0.963, Q^2=0.832, S=1.391, F=97.897$$

按式(3)给出的计算值与相应实验值(表 1)基本吻合。用变异膨胀因子(variance inflation factors, VIF)<sup>[13]</sup>评价模型(3)中各自变量之间是否存在多重相关性。VIF 的定义式为:

$$VIF=1/(1-R^2) \quad (4)$$

式中:  $R^2$  为自变量集中某一变量与余下变量的判定系数。如  $VIF=1$ , 表明各自变量间完全不相关;当  $VIF < 5$  时,说明变量间没有明显的自相关性,所建模型是稳定的;当  $VIF > 5$  时,说明变量间存在明显的共线性,所建模型不能用于估算与预测。模型(3)中  $^0X_p^v$ 、 $M_{82}$ 、 $M_{22}$ 、 $M_{14}$  的 VIF 依次为 2.076、5.043、4.548、1.725, 它们的 VIF 大都小于或非常接近 5, 证明该模型中变量间没有明显的自相关性,具有良好的稳健度及预测能力。另外,模型(3)的  $R_{adj}^2$  与  $Q^2$  相差为 0.121, 小于 0.3, 说明该模型没有过拟合、不存在不相关的其他变量或数据中存在离域点。

### 3 讨 论

化合物的色谱保留时间受固定相与溶质分子间的作用力控制,其间作用力越大,  $t_R$  越大。蚌兰花中所含 20 种挥发性化合物均为弱极性分子,它们与固定相之间的作用力包括取向力、诱导力、色散力和氢键,但以色散力为主<sup>[7]</sup>。影响色散力的主要结构因素是分子的大

小与空间形状。化合物分子的体积越大,其变形性越大,相应瞬间偶极越强,其与固定相之间的色散力越大。分子中所含碳原子支化度越大,分子中各个原子与固定相距离较近,其间作用力较大。因此,它们的色谱保留时间随之增长。根据进入模型(3)的分子描述符可知,  $^0X_p^v$  主要反映分子的大小,  $M_{14}$  (反映第二类原子间相互作用),  $M_{22}$  (反映第二类原子与第九类原子间相互作用),  $M_{82}$  (反映第十类原子与第十类原子间相互作用)则与分子形状相关。因此,式(3)的削减误差比例(即  $R^2$ ) 为 96.3%, 只有近 3.7% 的影响  $t_R$  的其他因素未被揭示。可以认为式(3)中的  $^0X_p^v$ 、 $M_{14}$ 、 $M_{22}$ 、 $M_{82}$  及常数项较好地表征了影响  $t_R$  的本质因素。

综上所述,所建模型具有良好的相关性与稳健性,表明连接性指数、电性距离矢量对蚌兰花中所含挥发性化学成分的分子结构表征是合理的,确实揭示了影响本质因素,并较好解释了  $t_R$  的递变规律。

### 参 考 文 献:

- [1] 黄丽莎, 朱峰. 蚌兰花挥发油化学成分的 GC-MS 分析[J]. 中药材, 2009, 32(1): 65-66.
- [2] 白成科, 张媛, 王喆. 莱菔子脂肪酸成分的 GC-MS 分析[J]. 热带亚热带植物学报, 2006, 14(5): 409-412.
- [3] RANDIC M. On characterization of molecular branching[J]. J Amer Chem Soc, 1975, 97(23): 6609-6615.
- [4] 陈艳, 堵锡华. 红莓发酵香气成分定量结构-色谱保留值构效关系研究[J]. 食品科学, 2009, 30(21): 39-42.
- [5] 冯长君. 用分子形状指数估算氯代烷基苯甲醛的气相色谱保留指数[J]. 化工科技, 2007, 15(4): 9-12.
- [6] 陈艳, 堵锡华. 天然山楂香料挥发性组分分子定量结构-色谱保留研究[J]. 食品科学, 2009, 30(18): 270-273.
- [7] 冯长君, 堵锡华. 胺类化合物 Kováts 指数的拓扑研究[J]. 色谱, 2001, 19(2): 124-127.
- [8] 郭宗儒. 药物分子设计[M]. 北京: 科学出版社, 2006: 477.
- [9] KIER L B, HALL L H. Molecular connectivity in structure-activity analysis[M]. England: Research Studies Press, 1986: 82.
- [10] LIU Shushen, YIN Chunsheng, WANG Liansheng. Combined MEDV-GA-MLR METHOD for QSAR of three panels of steroids, dipeptides, and COX-2 inhibitors[J]. J Chem Inf Comput Sci, 2002, 42: 749-756.
- [11] LIU Shushen, LIU Hailing, YIN Chunsheng, et al. VSMP: A novel variable selection and modeling method based on the prediction[J]. J Chem Inf Comput Sci, 2003, 43(8): 964-969.
- [12] LIU Shushen, YIN Chunsheng, LI Zhiliang, et al. QSAR study of steroid benchmark and dipeptides based on MEDV-13[J]. J Chem Inf Comput Sci, 2001, 41(2): 321-329.
- [13] 冯长君. 手性有机酸保留指数的手性指数及原子类型电拓扑指数模型[J]. 物理化学学报, 2010, 26(1): 193-198.