

储良龙眼挥发性成分的定量 - 保留指数分析

冯长君, 刘玉胜, 冯 惠
(徐州工程学院化学化工学院, 江苏 徐州 221008)

摘 要: 基于拓扑理论及 MATLAB 程序计算 41 种龙眼挥发性成分的分子连接性指数(mX_v)和电性距离矢量(M_i)。采用最佳变量子集回归方法, 建立龙眼挥发性成分的保留指数(B_z)与 mX_v 、 M_i 的定量结构 - 保留相关的五元回归方程, 传统判定系数、逐一剔除法交叉验证系数依次为 0.981、0.976。用 Jackknife 法检验模型的稳健性, 其传统判定系数在 0.975~0.986 之间。该模型具有高度稳健性与良好的预测能力, 并能较好解释有机组分保留指数的递变规律。
关键词: 储良龙眼; 挥发性成分; 保留指数; 连接性指数; 电性距离矢量; 定量结构 - 保留相关性

Quantitative Structure-Retention Relationship Analysis of Volatile Compounds in Longan

FENG Chang-jun, LIU Yu-sheng, FENG Hui
(School of Chemistry and Chemical Engineering, Xuzhou Institute of Technology, Xuzhou 221008, China)

Abstract: Based on the topological theory and self-MATLAB program, Kier's molecular connectivity indices (mX_v) and electro-negativity distance vector (M_i) were calculated for 41 volatile compounds in Chuliang Longan as identified by GC-MS. A five-element regression model of quantitative structure-retention relationship (QSRR) for retention index (RI) as a function of mX_v and M_i was constructed using leaps-and-bounds regression (LBR). The traditional correlation coefficient (R^2) and the cross-validation correlation coefficient (Q^2) of leave-one-out (LOO) were 0.981 and 0.976, respectively. The robustness of the regression model was validated by Jackknife method, and the results showed that the R^2 was between 0.975 and 0.986. The present study demonstrates that the model is highly reliable and has favorable predictive ability, and can better elucidate the change rule of GC-MS retention indices for organic components.

Key words: Chuliang longan; volatile composition; retention index; molecular connectivity indices; electro-negativity distance vector; quantitative structure-retention relationship (QSRR)

中图分类号: TS207.3; O6-051

文献标识码: A

文章编号: 1002-6630(2012)08-0244-04

龙眼是无患子科(Sapindaceae)龙眼属(*Dimocarpus* Lour.)常绿果树, 俗名桂圆, 又名龙目、比目等。我国是龙眼原产地, 种植面积及产量居世界第一位。龙眼具有药食两用性。《神农本草经》称龙眼“主五脏邪气, 安志、厌食、久服强魂魄, 聪明”。现代药理研究表明, 龙眼能提高人体的免疫功能, 具有抗衰老、抗肿瘤、抵御细菌病毒的侵袭及抗炎、抗感染等作用^[1]。龙眼营养丰富, 总糖含量高, 富含维生素、视黄醇、尼克酸等。此外, 还含有粗蛋白、无机盐等人体所必需的营养物质。因此, 人们已开发出龙眼系列产品, 如龙眼混浊饮料、龙眼保健茶、龙眼发酵酒、龙眼保健果醋等, 但对龙眼(储良)中的芳香物质进行定量构效关系研究^[3-5]尚未见报道。本实验采用 Kier 和

Hall 的价连接性指数(mX_v)^[6]及 Liu 等^[7-9]的电性距离矢量(M_i), 关联储良龙眼中 41 种挥发性化合物的保留指数(B_z)^[10], 经最佳变量子集回归(leaps-and-bounds regression, LBR)建立的五元数学模型, 其相关系数(R)、判定系数(R^2)、交叉验证相关系数(Q^2)依次为 0.991、0.981、0.976, 呈现出良好稳定性与预测能力, 可为各种龙眼挥发性化合物保留指数的预测提供一种简便有效手段。

1 材料与amp;方法

1.1 材料与仪器

储良龙眼 市购。气相色谱 - 质谱联用仪 美国 Agilent 公司。

收稿日期: 2011-04-04

基金项目: 国家自然科学基金面上项目(21075138); 徐州市科技局基金项目(XZZD1104);

徐州市贾汪区科技局基金项目(XM10A05)

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

1.2 方法

1.2.1 Kier 指数构建

进行物质构效关系研究的关键是建立描述分子结构的拓扑指数^[11-12],以反映分子中原子间的连接方式与次序,即分子结构差异。迄今已报道了 400 余种拓扑指数,其中以 Randic^[13]最先提出、后经 Kier 等^[6]扩展的价连接性指数(${}^mX^t_v$)的应用最为广泛。Kier 指数的核心概念为原子点价(δ_i^v),用以表征非氢原子的结构差异,定义式为:

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

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。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(${}^mX^t_v$):

$${}^mX^t_v = \sum (\delta_i^v \delta_j^v \dots)^{0.5} \quad (2)$$

式中: ${}^mX^t_v$ 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 代表分子子图的类型,常用 4 种子图分别为链、星、星-链、环状子图,依次对应 t 为 p 、 c 、 pc 及 ch 。本实验利用 MATLAB 软件计算 11 种 ${}^mX^t_v$: ${}^0X_p^v$ 、 ${}^1X_p^v$ 、 ${}^2X_p^v$ 、 ${}^3X_p^v$ 、 ${}^4X_p^v$ 、 ${}^5X_p^v$ 、 ${}^3X_c^v$ 、 ${}^4X_c^v$ 、 ${}^4X_{pc}^v$ 、 ${}^5X_{pc}^v$ 、 X_{ch}^v 。

Liu 等^[7-9]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 M_t ,其计算过程分为以下 3 步进行:

首先,定义分子中非氢原子的原子类型(K_i)与原子属性(I_i),电性距离矢量理论给出 13 种原子类型及 43 种原子属性;如卤素属第 13 种原子类型,其中氯原子的 $I_{Cl} = 1.9108$ 。

第 2 步:考虑非氢原子 i 在分子中并不是孤立存在的,还要受其他非氢原子(j)的影响(相应扰动值为“ ΔI_i ”),定义其电性状态指数(D_i)为:

$$D_i = I_i + \Delta I_i \quad (3)$$

第 3 步:综合考虑不同原子类型的非氢原子之间的电性及拓扑作用,定义分子电性距离矢量(molecular electronegativity distance vector, MEDV),以“ M_t ”表示。这 13 种原子类型两两相互作用,使构成 91 个“ M_t ”, $t=1\sim 91$ (具体计算过程见文献[7-9])。

本实验化合物分子中共有 6 种原子类型: $K_i=1,2,3,4,9,10$ 。它们两两组合构成的 M_t 的依次为 M_1 、 M_2 、 M_3 、 M_4 、 M_9 、 M_{10} 、 M_{14} 、 M_{15} 、 M_{16} 、 M_{21} 、 M_{22} 、 M_{26} 、 M_{27} 、 M_{32} 、 M_{37} 、 M_{42} 、 M_{43} 、 M_{77} 、 M_{78} 、 M_{82} 。本实验利用 MATLAB 程序计算了文献[10]中 41 种有机分子的这 21 个 M_t (其中 M_{43} 、 M_{82} 全为 0)。

1.2.2 多元线性回归分析

将每种化合物的 11 种连接性指数、19 种电性距离矢量及相应的保留指数(B_z)输入 MINITAB14 软件,应用其中最佳量子集回归选择最佳变量组合,建立相应的

定量结构-保留相关(quantitative structure-retention relationship, QSRR)模型。采用逐一剔除法(leave-one-out, LOO)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。一般认为 Q^2 大于 0.5,模型比较稳定,具有良好的预测能力;大于 0.9,模型非常稳定及更优的预测准确率。用方差膨胀因子(variance inflation factors, VIF)^[14-15]评价模型中各自变量的多重相关性,VIF 的定义式为:

$$VIF = \frac{1}{1 - R^2} \quad (4)$$

式中: R^2 为自变量 X 中某一变量与余下变量的相关系数。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。

2 结果与分析

2.1 龙眼中化学组分分析

采用顶空固相微萃取法提取龙眼中挥发性成分,用气相色谱-质谱联用仪进行检测得到 41 种挥发性化合物的保留指数(B_z)见表 1(使用 MATLAB、MINITAB14 等计算软件进行统计分析)^[10]。

2.2 龙眼挥发性成分的 B_z 与 M_t 的数学模型

应用 MINITAB14 软件中最佳量子集回归程序建立的最佳 QSRR 模型:

$$B_z = 128.214 + 116.144 {}^0X_p^v + 3.764 M_{14} + 10.816 M_{15} + 22.257 M_{21} + 62.180 M_{77} \quad (5)$$

式中: n' 、 R 、 R^2 、 Q^2 、 F 、 S 分别为样本数、相关系数、判定系数(亦称削减误差比例)、交叉验证相关系数、Fischer 检验值、估计标准误差, $n'=41$, $R=0.991$, $R^2=0.981$, $Q^2=0.976$, $F=371.319$, $S=45.894$ 。其 $Q^2=0.976>0.9$,表明该模型非常稳定,具有很优的预测能力。按此模型给出的计算值与其实验值非常吻合(表 1),其关系见图 1。

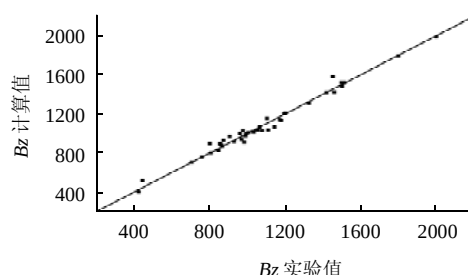


图 1 41 种龙眼挥发性化合物保留指数的实验值与计算值的关系
Fig.1 Relationship between experimental and calculated retention indexes of 41 volatile compounds in Chuliang longan

表 1 龙眼中 41 种化学组分的拓扑指数及其保留指数
Table 1 Molecular topological indices and GC-MS retention indices retention indices of 41 volatile compounds in Chuliang longan

序号	化学组分	$^0X_p^v$	M_{14}	M_{15}	M_{21}	M_{77}	B_z	
							实验值	计算值
1	甲苯 toluene	4.3868	24.1130	2.7484	0.0000	0.0000	759	758
2	辛烷 octane	6.2426	10.8730	0.0000	0.0000	0.0000	801	894
3	乙苯 ethylbenzene	5.0939	26.5370	5.8566	0.0000	0.0000	858	883
4	邻二甲苯 <i>p</i> -xylene	5.3094	17.9550	5.4002	0.0000	0.0000	866	871
5	1-乙基-3-甲苯 3-ethyltoluene	6.0165	17.2550	9.3774	0.0000	0.0000	959	993
6	β -月桂烯 β -myrcene	6.9831	3.3151	3.4420	0.0000	0.0000	992	989
7	柠檬烯 limonene	6.9831	5.3204	5.4207	0.0000	0.0000	1026	1018
8	反-罗勒烯 <i>trans</i> -ocimene	7.1463	4.4206	4.9993	0.0000	0.0000	1053	1029
9	γ -松油烯 γ -terpinene	7.1463	12.4110	4.9347	0.0000	0.0000	1063	1058
10	1,3,8- <i>p</i> -孟三烯 1,3,8- <i>p</i> -menthatriene	6.7760	10.2050	7.1672	0.0000	0.0000	1108	1031
11	别-罗勒烯 <i>allo</i> -ocimene	7.3094	14.1940	2.4707	0.0000	0.0000	1141	1057
12	萘 naphthalene	5.6188	39.8730	18.1690	0.0000	0.0000	1176	112
13	β -石竹烯 β -caryophyllene	9.3973	8.9432	14.8760	0.0000	0.0000	1416	1414
14	α -葎草烯 α -humulene	10.6380	22.6200	11.4730	0.0000	0.0000	1450	1573
15	β -法尼烯 β -farnesene	9.9747	11.5420	8.1126	0.0000	0.0000	1457	1418
16	<i>Z,E</i> - α -法尼烯 <i>Z,E</i> - α -farnesene	10.6380	9.7910	10.8230	0.0000	0.0000	1495	1518
17	α -古芸烯 α -gurjunene	10.6380	4.9534	8.8647	0.0000	0.0000	1499	1478
18	α -法尼烯 α -farnesene	10.6380	9.7910	10.8230	0.0000	0.0000	1508	1518
19	乙醇 ethanol	2.8614	0.3245	0.0000	2.2713	0.0000	445	512
20	正己醇 1-hexanol	5.6899	9.3647	0.0000	4.5173	0.0000	873	925
21	2-庚醇 2-heptanol	6.5601	8.3214	— 1.3500	2.6448	0.0000	905	966
22	正庚醇 1-heptanol	6.3970	12.4280	0.0000	4.6977	0.0000	975	1023
23	蘑菇醇 1-octen-3-ol	6.1374	5.7032	— 3.8377	3.9640	0.0000	982	909
24	2-乙基己醇 ethylhexanol	6.5601	4.6710	1.0200	4.1870	0.0000	1032	1012
25	正辛醇 1-octanol	6.3970	12.4280	0.0000	4.6977	0.0000	1075	1023
26	沉香醇 linalool	7.7232	3.8470	2.1685	3.9549	0.0000	1101	1151
27	丁酸乙酯 butanoic acid ethyl ester	5.4378	0.3989	— 0.2695	1.8481	0.0000	806	799
28	2-丁烯酸乙酯 2-butenic acid ethyl ester	5.1783	2.8268	— 0.7945	4.2076	0.0000	846	825
29	己酸甲酯 hexanoic acid methyl ester	6.1449	3.4785	— 0.3993	2.8403	0.0000	928	914
30	2-己烯酸甲酯 2-hexenoic acid methyl ester	5.8854	7.2613	— 0.9912	4.9357	0.0000	968	938
31	己酸乙酯 hexanoic acid ethyl ester	6.8520	3.6347	— 0.3385	3.2707	0.0000	1003	1007
32	安息香酸乙酯 benzoic acid ethyl ester	6.4104	22.7190	6.6757	4.8831	0.0000	1170	1139
33	水杨酸甲酯 2-hydroxybenzoic acid methyl ester	6.0731	15.9380	4.4036	7.6720	1.3236	1192	1194
34	辛酸乙酯 octanoic acid ethyl ester	8.2662	8.7739	— 0.3550	3.9737	0.0000	1199	1206
35	甲基香叶草酯 methyl geraniate	8.8854	4.4834	1.7862	4.8760	0.0000	1324	1305
36	十四碳酸乙酯 tetradecanoic acid ethyl ester	12.5090	27.7890	— 0.3672	4.8305	0.0000	1795	1789
37	十六碳酸乙酯 hexadecanoic acid ethyl ester	13.9230	34.5910	— 0.3685	4.9665	0.0000	1996	1982
38	乙醛 acetaldehyde	1.9856	0.0000	0.0000	1.8789	0.0000	425	401
39	<i>E</i> -2-己烯醛 <i>E</i> -2-hexenal	4.5545	9.3130	0.0000	8.9215	0.0000	854	891
40	乙酸 acetic acid	2.3555	0.0000	0.0000	0.0000	4.8665	705	704
41	3-辛酮 3-octanone	6.4438	3.8629	— 0.0485	3.5898	0.0000	988	971

2.3 所建数学模型的鲁棒性检验

为了检验模型(5)中是否存在“异常值”及机会相关,采用 Jackknife 法^[16]予以检验。其一,由于所研究的是较大样本(样本容量 $n' > 30$),采用逐组剔除法,即每次分别剔除化合物序号的个位是 1,2,⋯,0 的化合物,用余下数据建模,共进行 10 次,它们的 R^2 、 F 值及 S 值见表 2。

表 2 模型(5)的稳健性检验
Table 2 Results of determination of the robustness of model 5 by Jackknife method

剔除序号	1	2	3	4	5	6	7	8	9	0
R^2	0.984	0.986	0.982	0.985	0.982	0.978	0.975	0.978	0.981	0.983
F	361.8	447.4	344.2	413.8	339.6	281.3	237.7	280.5	320.3	360.5
S	45.60	41.35	46.57	42.44	46.63	47.61	48.43	48.07	46.18	45.39

由表2可见, 10个模型的 R^2 在0.975~0.986之间, 即在0.981附近上下波动, 呈现良好的正态分布。说明模型(5)对41个化合物的保留指数具有良好稳定性。模型(5)中各自变量 $^0X_p^v$ 、 M_{14} 、 M_{15} 、 M_{21} 、 M_{77} 的VIF值依次为1.276、1.354、1.976、1.583、1.112, 它们的VIF都远小于5, 可见各个自变量之间的自相关性很低, 再次证明该模型具有较好的稳定性。

3 讨论与结论

在气相色谱过程中, 除了溶质本性以外的所有影响因素都可人为予以有效控制, 因此, 溶质的性质便成为该过程中的唯一变数。显然, 气相色谱保留指数(B_z)便是溶质分子微观结构的函数。因此, 通过物质的定量结构与 B_z 的QSRR, 不仅可以显示分子结构的微观参数与 B_z 的递变规律, 而且对于预测保留值、选择分离条件以及探索色谱保留机理具有重要的意义。在柱温及固定相确定情况下, B_z 则只与溶质的分子结构相关, 即与溶质分子间的作用能有关。分子间作用能的本质是分子间力——取向力、诱导力、色散力和氢键, 通常以色散力(或氢键)为主。影响色散力的主要因素为分子的大小与形状: 溶质分子的体积越大, 其变形性越大, 相应瞬间偶极之间的作用力增强, 其色散力越大, 相应它们的 B_z 越大。对于同分异构体, 其色散力基本上是随着分子的分枝度增大而减小的。因此, 它们的气相色谱保留指数与分子的分枝度呈负相关。本实验所研究的化合物有烷烃、芳烃、醇、有机酸、酯等, 其分子之间含有取向力、诱导力、色散力及氢键等。进入模型(5)中的自变量为 $^0X_p^v$ 、 M_{14} 、 M_{15} 、 M_{21} 、 M_{77} , 它们所蕴含的物理意义如下: $^0X_p^v$ 反映分子大小及其中非氢原子结构差异; 电性距离矢量 $M_{14}(2 \times 2)$ 、 $M_{15}(2 \times 3)$ 、 $M_{21}(2 \times 9)$ 、 $M_{77}(9 \times 9)$ 涉及第2类原子“—C—”(非极性基团)、第3类原子“—C<”(非极性基团)及第9类原子“—O”(强极性基团, 即羟基)之间的相互作用, 其间包含色散力、取向力、氢键等作用力。因此, 这些描述子与保留指数之间存在密切相关关系。模型(5)的相关系数高达0.991, 这在食品科学中的QSRR研究已属少见^[3,17-19], 这也表明其对龙眼挥发性成分的结构表征是合理的, 为龙眼的综合利用以及功能性食品的研发提供理论支持。

采用价连接性指数($^mX_p^v$)、电性距离矢量(M_i)与储良龙眼中挥发性化合物的保留指数(B_z)关联, 经最佳变量子集回归建立的数学模型, 具有较强的稳健性及预测能力, 其削减误差比例(R^2)高达98.1%, 说明仅有1.9%的随机因素尚未被揭示。根据进入模型(5)中的自变量

$^0X_p^v$ 、 M_{14} 、 M_{15} 、 M_{21} 、 M_{77} 可知, 影响保留指数的结构因素主要是一C—、—C<、羟基及其相互之间的作用力。利用所建的 B_z -QSRR模型, 不仅可以解释 B_z 的递变规律, 而且可以估算与预测保留值, 对于选择实验分离条件、探索色谱保留机理等也具有一定的参考价值。

参考文献:

- [1] 李升锋, 刘学铭, 吴继军, 等. 龙眼果肉的研究与开发[J]. 福建果树, 2004, 25(2): 12-15.
- [2] 文良娟, 滕建文, 于兰. 龙眼果汁饮料的研制[J]. 食品科技, 2002, 28(9): 55-56.
- [3] 堵锡华. 香梨酒香气成分保留时间的定量构效关系研究[J]. 食品科学, 2011, 32(2): 218-221.
- [4] 余训民, 杨道武. 应用新定义的拓扑指数预测烷氧氯硅烷、单硫醚的气相色谱保留指数[J]. 分析化学, 2005, 33(1): 101-105.
- [5] FENG Changjun, YANG Weihua, MU Lailong. Estimation and prediction of bioconcentration factors of nonionic organic chemicals in fish by electrotopological state indices and structural parameter[J]. Chinese Journal of Structural Chemistry, 2008, 27(5): 575-587.
- [6] KIER L B, HALL L H. Derivation and significance of valence molecular connectivity[J]. Journal of Pharmaceutical Sciences, 1981, 70(6): 583-589.
- [7] LIU Shushen, YIN Chunsheng, LI Zhiliang, et al. QSAR study of steroid benchmark and dipeptides based on MEDV-13[J]. Journal of Chemical Information and Computer Sciences, 2001, 41(2): 321-329.
- [8] LIU Shushen, LIU Hailing, YIN Chunsheng, et al. VSMP: A novel variable selection and modeling method based on the prediction[J]. Journal of Chemical Information and Computer Sciences, 2003, 43(3): 964-969.
- [9] 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(3): 749-756.
- [10] 张义, 高蓓, 徐玉娟, 等. 顶空固相微萃取-气质联用方法分析龙眼中的挥发性化合物[J]. 食品科学, 2010, 31(16): 156-160.
- [11] 冯长君. 取代芳烃生物降解性的新型连接性指数模型[J]. 华中科技大学学报: 自然科学版, 2010, 38(3): 108-111.
- [12] 冯长君, 沐来龙, 杨伟华, 等. 用拓扑指数和神经网络研究有机污染物的生物富集因子[J]. 化学学报, 2008, 66(19): 2093-2098.
- [13] RANDIC M. On characterization of molecular branching[J]. J Amer Chem Soc, 1975, 97(23): 6609-6615.
- [14] 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.
- [15] 冯长君. 3-取代巯基-5-(2-羟基苯基)-4H-1,2,4-三唑类化合物抑菌活性的定量构效关系和结构修饰的理论研究[J]. 化学学报, 2012, 70(4): 512-518.
- [16] 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.
- [17] 堵锡华, 陈艳. 柚子皮香精油挥发性成分的保留相关性研究[J]. 食品科学, 2009, 30(19): 61-64.
- [18] 陈艳, 堵锡华. 红莓发酵酒香气成分定量结构-色谱保留值构效关系研究[J]. 食品科学, 2009, 30(21): 39-42.
- [19] 冯长君, 石春玲, 李鸣建. 蚌兰花挥发油化学成分色谱保留值的构效关系研究[J]. 食品科学, 2010, 31(19): 42-44.