

碳氢化合物、醛、酮和硫醇的定量结构 - 色谱保留指数相关性研究

周丛艺¹, 聂长明¹, 田万福², 戴益民³, 彭国文¹

(1. 南华大学 化学化工学院, 湖南 衡阳 421001; 2 湘南中学, 湖南 郴州 423000;
3 长沙理工大学 化学与环境工程系, 湖南 长沙 410077)

摘要: 采用平衡电负性和相对化学键长对传统距离矩阵进行修正, 构建新拓扑指数 N_t 。结合路径数, 建立碳氢化合物、醛、酮和硫醇等化合物在 24 种极性和非极性色谱柱上的定量结构 - 色谱保留指数关系 (QSRR) 模型, 23 种模型的相关系数大于 0.99。模型经留 n 法交叉检验, 显示出良好稳健性和预测能力。模型物理意义明确, 表明色谱保留指数可用分子的大小、平衡电负性、支化度和形状等内在结构信息进行有效表征。模型经 Needham 公式分析, 结果显示新指数 N_t 对保留指数影响最大。借助 Hyperchem 软件进行对比研究, 结果表明拓扑化学法优于量子化学 AM1 法。

关键词: 拓扑指数 (N_t); 定量结构 - 色谱保留指数关系 (QSRR); 碳氢化合物; 醛; 酮; 硫醇

中图分类号: O6-04 **文献标识码:** A **文章编号:** 1004-4957(2007)06-0802-06

Study on the Quantitative Structure - Chromatographic Retention Indices Relationships of Hydrocarbons, Aldehydes, Ketones and Mercaptans

ZHOU Chong-yi¹, NIE Chang-ming¹, TIAN Wan-fu², DAI Yi-min³, PENG Guo-wen¹

(1. School of Chemistry and Chemical Engineering, Nanhua University, Hengyang 421001, China; 2. Xiangnan Middle School, Chenzhou 423000, China; 3. Department of Chemistry and Environmental Engineering, Changsha University of Science and Technology, Changsha 410077, China)

Abstract: A novel topological index N_t was established by using the equilibrium electronegativity and the relative chemical bond length to modify the traditional distance matrix. Along with the path number, the models of the quantitative structures - chromatographic retention indices relationships (QSRR) of hydrocarbons, aldehydes, ketones and mercaptans on 24 stationary phases including the polar and nonpolar ones were established. It was found that the correlation coefficients of 23 models were all greater than 0.99. The leave- n -out cross validation indicated that the models were statistically significant and reliable. The physical meaning of the model was explicit indicating that the chromatographic retention index (RI) could be characterized efficiently by the size and shape of the molecules, equilibrium electronegativity and branching degree. The results obtained by analyzing the model with the Needham equation demonstrated that the new index N_t has the greatest impact on the retention index. The results obtained from the comparison study with the Hyperchem software also indicated that the topological chemical method was superior to the AM1 method of quantum chemistry.

Key words: Topological index (N_t); Quantitative structure - retention relationships (QSRR); Hydrocarbon; Aldehyde; Ketone; Mercaptan

作为色谱科学的一个研究分支, 定量结构 - 色谱保留指数关系 QSRR (quantitative structure - retention relationship) 的研究和应用已获得越来越多科研工作者的关注。目前 QSRR 主要应用于以下方面^[1-2]: 预测新溶质的保留; 确定溶质最具信息量的结构描述符; 探讨在一个已有色谱体系中分离操作的分子机理; 评价分析物的复杂物化性质; 预测相关的生物活性。

经过几十年发展, 现在 QSRR 研究主要有 3 种方法。最早 1 种是用正辛醇 - 水分配系数的对数 ($\log P$) 来关联保留指数^[3]; 第 2 种是借助线性溶剂化能相关 (linear solvation energy relationships) 方法进

收稿日期: 2006-10-27; 修回日期: 2006-11-18

基金项目: 湖南省自然科学基金资助项目 (03JJY3024); 湖南省经委技术创新基金资助项目 (湘经科 [2005] 283 号)

作者简介: 周丛艺 (1978-), 男, 湖南郴州人, 硕士研究生; 通讯作者: 聂长明, Tel: 0734-8282057, E-mail: niecm196132@163.com

行研究^[4-5]; 第 3 种是从计算化学角度出发, 借助各种拓扑指数, 电子参数, 几何参数以及物理参数等结构描述符来建立 QSRR 模型, 该方法已成为 QSRR 研究中的热点^[6-7]。

虽然上述研究都取得较好结果, 但利用同一方法应用于多样本数化合物在不同色谱柱上的气相色谱保留指数 (以下用 R_I 表示) 研究并不多见。本文借助平衡电负性和相对化学键长对传统距离矩阵进行修正, 构建新拓扑指数 N_t 。用该指数和路径数对碳氢化合物、醛、酮和硫醇等化合物 24 种极性和非极性色谱柱上的定量结构 - 色谱保留指数关系 (QSRR) 模型进行研究。同时, 利用量子化学法进行对比研究。研究表明: 本文所构建的拓扑指数 N_t 和 QSRR 模型, 能够合理有效地表征化合物的 R_I 变化规律, 其结果在应用范围及相关程度两个层面上都取得了优于文献的量子化学 AM1 方法。

1 原理与方法

1.1 拓扑指数 N_t

距离矩阵可表示为 $D = (d_{ij})_{n \times n}$, 其中 n 为分子中的非氢原子个数, d_{ij} 为分子结构中基团 i 和 j 之间的最短拓扑边长之和。传统的距离矩阵视两顶点 i 和 j 的最小距离为 1, 为更合理有效表征化合物分子的几何本质结构特征, 本文采用化学键长与碳碳单键键长的比值作为矩阵中最小边长。表 1 列出了一些化学键类型和相对化学键长^[10]。

表 1 一些化学键类型和相对化学键长 (L/L_{C-C})
Table 1 Some types of chemical bonds and their relative chemical bond length (L/L_{C-C})

Bond	L/L_{C-C}	Bond	L/L_{C-C}	Bond	L/L_{C-C}	Bond	L/L_{C-C}
$>C-C<$	0.154/0.154	$\geq C-C \leq$	0.148/0.154	$\equiv C-C \equiv$	0.138/0.154	$>C=C<$	0.134/0.154
$=C=C=$	0.128/0.154	$-C \equiv C-$	0.121/0.154	$C-S$	0.182/0.154	$C-Cl$	0.177/0.154
$C-Br$	0.193/0.154	$C-I$	0.214/0.154	$Si-Cl$	0.202/0.154	$Si-O$	0.153/0.154

电负性是描述原子性质的重要参数之一, 主要是表示原子得失电子的能力, 而平衡电负性可以表征化合物中同种原子在不同化学环境中的电负性不同的特点^[11]: $\chi_i = (\chi_A + \sum G) / (1 + \sum D)$, 式中 χ_A 为原子 i 的鲍林电负性, $\sum D$ 表示与原子 i 直接相连的基团数之和, $\sum G$ 表示与原子 i 直接相连基团电负性之和, G 按下式进行计算^[11]: $G = \{ (1/n_{1l}) \sum_{l=1}^{n_{1l}} \{ (1/n_{2l}) \sum_{l=1}^{n_{2l}} \dots [(1/n_{kl}) \sum_{l=1}^{n_{kl}}] \dots \} \}$ 。

为有效糅合电性信息, 本实验把分子中顶点原子的平衡电负性加入到修正的距离矩阵的对角线上, 从而得到矩阵 D_1 :

$$D_1 = \begin{bmatrix} 1 & L_{12}/L_{C-C} & \dots & L_{1(n-1)}/L_{C-C} & L_{1n}/L_{C-C} \\ L_{21}/L_{C-C} & 2 & \dots & L_{2(n-1)}/L_{C-C} & L_{2n}/L_{C-C} \\ \dots & \dots & \dots & \dots & \dots \\ L_{i1}/L_{C-C} & \dots & i & \dots & L_{in}/L_{C-C} \\ \dots & \dots & \dots & \dots & \dots \\ L_{n1}/L_{C-C} & L_{n2}/L_{C-C} & \dots & L_{n(n-1)}/L_{C-C} & n \end{bmatrix}$$

然后定义新拓扑指数 N_t : $N_t = \lg N \lg \max$, 其中 N 为分子中的非氢原子个数, \max 为修正矩阵 D_1 的最大特征值。例如, 3-甲基-1-丁硫醇的原子骨架示意图见图 1。

根据上述方法, 其 $N_t = \lg N \lg \max = \lg 6 \lg 13.7373 = 0.8855$ 。

1.2 路径数

路径数 (P_n) 由 Wiener^[12] 率先提出, 主要是指两原子间间隔若干个化学单键的数目 (即相距的边数 d_{ij})。Gordon 与 Scantlebury^[13] 于 1964 年提出 P_2 , 用以反映分子支化度信息。 P_3 由 Wiener^[12] 提出, 称为极化数, 用以

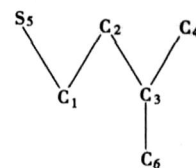


图 1 3-甲基-1-丁硫醇原子骨架示意图
Fig. 1 The atom skeletal description of 3-methyl-1-thiol

表征分子的形状信息。在路径数选择上,采用主成分分析(PCA)法对描述符 P_n 进行信息量压缩和选择分析,最后挑选 P_2 和 P_3 来进行建模。尽管有些学者已对烷烃异构体的路径数进行了研究^[14-15],但是关于具有多样性组成的 R_I 研究较少,本文所述方法正推广应用于较复杂的杂原子化合物,取得满意效果。

2 结果与讨论

2.1 模型建立

鉴于指数 N_I 主要包涵分子的大小和电负性等结构信息, P_2 和 P_3 体现分子支化度和形状等结构信息;而化合物的 R_I 与分子的结构有密切联系,因而本文采用多元线性回归(MLR)技术建立模型^[16]: Property 为 $a_0 + a_1 N_I + a_2 P_2 + a_3 P_3$, 在建立烷烃的 QSRR 模型时,参照文献[9],对 R_I 进行对数变换。由于烷烃为非极性分子,分子中主要为色散力;对于其它化合物,考虑到其分子作用力情况复杂些,建模时,不做对数变换。建模遵循 R 最大和 s 最小的基本原则;模型评价通过相关系数 R , 标准误差 s 和检验值 F 等参数实现。本文中烷烃在 12 种不同状况的固定相上的 R_I 来自文献[17-18];烷基苯 4 种 R_I 来自文献[19];醛酮 4 种 R_I 来自文献[19];硫醇 4 种 R_I 来自文献[18]。

2.1.1 烷烃的 QSRR 模型 首先对烷烃在 6 种不同固定相上的 R_I 建立模型。从表 2 的统计结果可知,本文所建立的 6 种烷烃模型 R 均大于 0.99,属于优良级别。其中有 3 种模型的 R 均大于 0.999 0,结果明显优于文献方法。在关联 BSA 上的 R_I 时,参照统计法 Dixon 准则对乙烷进行剔除^[20]。

表 2 烷烃回归模型统计结果

Table 2 The statistical results of regression model of alkanes

Stationary phase	$\lg R_I = a_0 + a_1 N_I + a_2 P_2 + a_3 P_3$				F	S	R	R_{CV}	$PRESS$	SSY	$PRESS/SSY$
	a_0	a_1	a_2	a_3							
SQ	-1.895 3	2.630 8	-0.050 9	0.071 1	5 327	0.025 1	0.999 5	0.996 7	1.296	196	0.006 6
TAS	-1.872 8	2.583 3	-0.050 0	0.075 7	5 213	0.025 4	0.999 5	0.996 6	1.370	202	0.006 8
DMS	-1.587 3	2.190 8	-0.051 4	0.076 9	2 923	0.029 1	0.999 0	0.995 0	1.036	105	0.009 9
ODP	-1.382 1	1.872 9	-0.041 3	0.065 0	925	0.044 5	0.997 0	0.995 0	0.534	53	0.010 1
DOP	-1.890 7	2.627 5	-0.053 1	0.068 4	2 282	0.037 9	0.998 8	0.996 3	1.220	168	0.007 3
BSA	-1.645 7	2.310 5	-0.061 9	0.078 3	667	0.052 4	0.996 0	0.990 1	1.967	100	0.019 7

另一方面,为考察该方法的适用性,实验把研究样本扩展到 49 种烷烃。对其在 4 种固定相上(其中 SQUALANE 又包含 30、50 和 70 3 种温度)共 6 种 R_I 进行研究。从表 3 统计结果可知, R 均大于 0.995,模型都属于优良级别。从表 2 和表 3 可看出, a_1 和 a_3 均为正值,而 a_2 都为负值。这说明烷烃 R_I 与指数 N_I 和 P_3 成正相关,与 P_2 成负相关。

表 3 烷烃回归模型统计结果

Table 3 The statistical results of regression model of alkanes

Stationary phase	$\lg R_I = a_0 + a_1 N_I + a_2 P_2 + a_3 P_3$				F	S	R	R_{CV}	$PRESS$	SSY	$PRESS/SSY$
	a_0	a_1	a_2	a_3							
JW DB-1	2.486 4	0.311 1	-0.005 1	0.009 2	1 959	0.004 6	0.996 2	0.995 5	3 143	351 476	0.008 9
H-P PONA	2.486 6	0.311 1	-0.005 1	0.009 2	1 951	0.004 7	0.996 2	0.995 5	3 153	351 337	0.009 0
JW DB-5	2.487 1	0.309 1	-0.005 1	0.009 5	1 936	0.004 7	0.996 1	0.995 5	3 202	353 854	0.009 0
SQUALANE ¹ *	2.490 0	0.307 3	-0.005 4	0.009 5	1 727	0.004 9	0.995 7	0.994 8	3 557	345 197	0.010 3
SQUALANE ² *	2.492 2	0.302 9	-0.005 1	0.009 7	1 688	0.005 0	0.995 6	0.994 7	3 677	348 377	0.010 6
SQUALANE ³ *	2.494 2	0.298 7	-0.004 8	0.010 0	1 650	0.005 1	0.995 5	0.994 6	3 791	352 002	0.010 8

* SQUALANE¹, SQUALANE², SQUALANE³ represent the temperature of 30, 50, 70, respectively

2.1.2 烷基苯的 QSRR 模型 相对而言,环状化合物的 QSRR 研究较链状化合物少些。因而实验对烷基苯在 4 种固定相上的 R_I 进行研究。从表 4 统计结果可知,模型都属于优良级别。 a_1 和 a_3 均为正值,这说明烷基苯 R_I 与分子的大小和形状有密切联系;而 a_2 则无固定规律。

表 4 烷基苯回归模型统计结果

Table 4 The statistical results of regression model of alkyl benzenes

Stationary phase	$\lg RI = a_0 + a_1 Nt + a_2 P_2 + a_3 P_3$				F	S	R	R_{CV}	PRESS	SSY	PRESS/SSY
	a_0	a_1	a_2	a_3							
SQUALANE	138.534 1	401.285 9	1.771 2	24.365 7	1 788	12 021 9	0.998 0	0.997 9	3 232	778 378	0.004 2
Ucon LB 550X	245.368 9	387.235 3	-4.616 7	31.248 1	1 292	14 412 8	0.997 3	0.997 1	4 614	809 765	0.005 7
Ucon HB 280X	310.272 2	371.023 7	-6.441 9	34.161 0	1 174	15 245 4	0.997 0	0.996 9	5 100	823 261	0.006 2
Carbowax 20M	458.854 8	320.115 9	-10.787 3	41.075 5	893	17 595 3	0.996 1	0.996 0	6 712	835 746	0.008 0

2.1.3 醛酮的 QSRR 模型 由于实际应用中研究对象常为含杂原子的化合物体系, 因此有必要验证该方法对烷烃之外的物质的适用性。首先对醛酮化合物在 4 种不同固定相上的 R 进行研究。从表 5 的统计结果可知, 该方法对醛酮也有较好的应用。4 种固定相极性强弱顺序为 HP-1 < HP-50 < DB-210 < Innovax^[19], 与其模型的回归相关度高低顺序相反。其中原因可能是随着固定相极性的增强, 溶质和固定相之间的作用力增加, 其作用机理更加复杂, 而拓扑模型揭示 R 的变化规律的能力也随之减弱。从表 5 可知, 方程中 a_1 均为正值, 仍说明醛酮 R 和指数 Nt 成正相关, 而 a_2 和 a_3 的数值正负则无一定规律。

表 5 醛酮回归模型统计结果

Table 5 The statistical results of regression model of aldehydes and ketones

Stationary phase	$\lg RI = a_0 + a_1 Nt + a_2 P_2 + a_3 P_3$				F	S	R	R_{CV}	PRESS	SSY	PRESS/SSY
	a_0	a_1	a_2	a_3							
HP-1	169.215 9	601.000 2	-6.856 3	1.109 2	1 911	11.032 5	0.998 0	0.997 9	2 874	700 523	0.004 1
HP-50	291.462 5	632.802 1	-8.373 4	-2.813 7	1 006	15.032 5	0.996 2	0.995 8	5 743	687 417	0.008 4
DB-210	414.891 1	613.024 4	12.147 0	-16.703 0	541	21.005 5	0.993 0	0.992 6	10 663	726 437	0.014 7
Innovax	483.464 8	678.476 8	-7.617 4	-16.322 3	323	24.380 7	0.988 3	0.987 9	14 155	590 018	0.024 0

2.1.4 硫醇的 QSRR 模型 硫醇的 QSPR 研究报道并不多见, 因而另一个研究体系采用了硫醇在 4 个固定相上的 R 。从表 6 的回归结果可看出, 该方法对硫醇的 R 也有较好的应用, R 均大于 0.997, 属于优良级别。模型中 a_1 和 a_3 均为正值, 而 a_2 都为负值。这说明硫醇 R 与指数 Nt 和 P_3 成正相关, 与 P_2 成负相关。

表 6 硫醇回归模型统计结果

Table 6 The statistical results of regression model of mercaptans

Stationary phase	$\lg RI = a_0 + a_1 Nt + a_2 P_2 + a_3 P_3$				F	S	R	R_{CV}	PRESS	SSY	PRESS/SSY
	a_0	a_1	a_2	a_3							
Apiezon M	299.844 7	606.647 0	-20.496 0	21.588 4	1 499	19.415 3	0.998 2	0.998 1	6 223	1 700 900	0.003 7
OV-17	380.381 9	614.548 7	-24.792 0	24.632 7	1 390	20.332 5	0.998 1	0.998 0	6 822	1 730 033	0.003 9
TritonX 305	472.291 9	638.502 9	-29.443 4	26.005 8	1 199	22.200 7	0.997 8	0.997 7	8 137	1 780 741	0.004 6
PEG-1000	524.710 9	662.021 7	-35.277 4	27.710 0	1 164	22.704 0	0.997 7	0.997 6	8 597	1 808 759	0.004 8

2.2 模型检验

为考察模型的稳健性和预测能力, 有必要进行模型检验。考虑到本文研究的样本数比较大, 因而采用留 n 交叉检验法 (leave- n -out cross-validation)^[21-22]。作为示例, 表中所列均为其中一次抽样建模结果。其中模型的稳健性和预测能力通过以下参数进行考察 R_{CV} 、PRESS 和 SSY ^[23]: 一般说来, 稳定的模型在交叉检验前的 R 和检验后的 R_{CV} 数值是比较接近的。同时, 如果 PRESS 远远小于 SSY 数值, 可证实模型具有良好的预测能力。而且, PRESS/SSY 可用来估计各模型的置信区间。如果 PRESS/SSY < 0.6, 则说明该模型可以接受; 如果 PRESS/SSY < 0.1, 则说明该模型属于优良级别^[23]。从表 2~5 的检验结果可看出, 本文所构建模型的 R 和 R_{CV} 数值比较接近, 而且 PRESS/SSY 介于 0.003 7 和 0.024 0, 均小于 0.1, 证明模型具有良好的稳健性和预测能力。

2.3 各指数对色谱保留指数的贡献

虽本文模型都符合建模要求, 但有必要讨论 3 个指数各自对整个模型所起的贡献, 进而探讨具有不

同物理意义的参数对 R_I 的不同影响。本文采用 Needham 等人所用公式^[24], 其中 r_i 表示相应指数对 R_I 的相对贡献, f_i 表示相应指数对 R_I 的百分数贡献, 二者定义如下: $r_i(i) = a_i T_i / \sum_i a_i T_i$, $f_i(i) = r_i^2 / \sum_i r_i^2 \times 100\%$, 式中 a_i 和 T_i 分别是方程 (4) 中第 i 个拓扑指数前的系数及拓扑指数平均值, r^2 为模型的判定系数。结果见表 7。

表 7 各化合物气相色谱指数模型的 r_i 和 f_i
Table 7 r_i and f_i of R_I models for the compounds

Compounds	Stationary phase	N_I	P_2	P_3
		$r_i(f_i)$	$r_i(f_i)$	$r_i(f_i)$
Alkanes(烷烃)	SQ	2.15 (81.36%)	-0.26 (9.84%)	0.23 (8.70%)
	TAS	2.11 (80.76%)	-0.26 (9.95%)	0.24 (9.19%)
	DMS	1.79 (77.67%)	-0.26 (11.28%)	0.25 (10.85%)
	ODP	1.53 (77.99%)	-0.21 (10.70%)	0.21 (10.70%)
	DOP	2.15 (81.24%)	-0.27 (10.20%)	0.22 (8.31%)
	BSA	1.89 (76.21%)	-0.32 (12.90%)	0.25 (10.08%)
	JW DB-1	0.37 (76.50%)	-0.05 (10.34%)	0.06 (12.41%)
	H-P PONA	0.37 (76.50%)	-0.05 (10.34%)	0.06 (12.41%)
	JW DB-5	0.37 (74.92%)	-0.05 (10.12%)	0.07 (14.17%)
	SQUALANE ¹	0.37 (74.86%)	-0.05 (10.12%)	0.07 (14.16%)
Alkyl benzenes (烷基苯)	SQUALANE ²	0.36 (74.34%)	-0.05 (10.33%)	0.07 (14.46%)
	SQUALANE ³	0.36 (75.91%)	-0.04 (8.43%)	0.07 (14.76%)
	SQUALANE	550.30 (61.25%)	21.04 (2.34%)	323.58 (36.01%)
Aldehydes and ketones(醛酮)	Ucon LB 550X	531.03 (52.77%)	-54.85 (5.45%)	414.97 (41.24%)
	Ucon HB 280X	508.8 (48.68%)	-76.53 (7.32%)	453.66 (43.40%)
	Carbowax20M	438.99 (39.15%)	-128.15 (11.43%)	545.48 (48.64%)
Mercaptans(硫醇)	HP-1	598.83 (92.76%)	-39.36 (6.10%)	4.81 (0.75%)
	HP-50	630.52 (90.66%)	-48.07 (6.91%)	-12.19 (1.75%)
	DB-210	610.82 (79.99%)	69.73 (9.13%)	-72.38 (9.48%)
	Innowax	676.03 (83.53%)	-43.73 (5.40%)	-70.73 (8.74%)
	Apiezon M	639.39 (75.81%)	-116.83 (13.85%)	84.19 (9.98%)
	OV-17	647.72 (72.90%)	-141.31 (15.90%)	96.07 (10.81%)
	TritonX 305	672.96 (71.11%)	-167.83 (17.73%)	101.42 (10.72%)
	PEG-1000	697.75 (68.98%)	-201.08 (19.88%)	108.07 (10.68%)

从表 7 可知, 3 个指数对所构建 QSRR 模型的贡献是不一样的。整体而言, N_I 的贡献较大; 而 P_2 和 P_3 的贡献较小, 但二者相差不是很大。3 指数物理意义明确, 主要体现了分子大小和电负性, 支化度和形状等内在结构信息。指数 N_I 与 R_I 良好的相关性与其体现化合物分子的内在结构信息是密不可分的: 非氢原子个数越多, 则 N_I 越大。如丙酮, 2-丁酮, 2-戊酮和 2-己酮的 N_I 分别为 0.504 8、0.687 6、0.871 5、1.050 4。平衡电负性客观描述分子中每个顶点原子的电负性特征, 从而真实反映分子中原子所处化学环境。例如, 乙酸丙酯中 7 个顶点原子的平衡电负性分别是: 2.421 2、2.788 8、2.838 0、2.471 8、2.349 9、2.303 1 和 3.005 8, 这与其各原子的化学环境各不相同的情况是一致的。相对传统距离矩阵中两顶点 i 和 j 的最小距离为 1 而言, 采用相对化学键长能更合理有效表征化合物分子的几何本质结构特征。指数 N_I 体现了分子中成键原子间的相互作用力的情况。如, 在醛酮化合物分子中, 羰基中的碳原子为 sp^2 杂化轨道, 与非羰基中的碳原子杂化轨道类型有所不同。该信息可体现在: 分子中不同成键类型的碳原子, 其平衡电负性的数值是不同的。研究显示本文拓扑法比文献的量子化学 AM1 方法可获得更好结果^[12], 本文模型在整体上体现了量子化学 AM1 方法所不具备的优势。

3 结 论

(1) 平衡电负性能够客观地反映原子在分子中所处的化学环境。采用相对化学键长作为距离矩阵中最小单位拓扑距离, 既遵循拓扑图论基本原理, 又合理反映分子几何结构特点。

(2) N_I 结合 P_2 和 P_3 所构建的 QSRR 模型能够合理有效的揭示碳氢化合物、醛酮和硫醇等化合物的不同气相色谱保留指数 (R_I) 变化规律。采用留 n 交叉法对模型进行检验, 参数 R_{CV} 、PRESS、SSY 和 PRESS/SSY 显示出模型良好稳健性和预测能力。通过 Needham 公式的指数贡献分析可知, N_I 有效体现

了决定 R_f 的主要因素的信息, 而路径数可起一个很好的补充作用。

(3) 用量子化学中的 AM1 方法扩展文献研究对象 (文献只研究烃类), 借助 Hyperchem 获取量子参数的方法更方便 (文献先用 Chemoffice 构建分子, 再用 MOPAC 计算参数)。比较表明, 本文所构建的新拓扑指数 N 和 QSRR 模型结果明显优于量子化学中的 AM1 方法。

参考文献:

- [1] KALISZAN R, VAN STRATEN M A, MARKUSZEWSKIM, et al. Molecular mechanism of retention in reversed-phase high performance liquid chromatography and classification of modern stationary phases by using quantitative structure-retention relationships[J]. *J Chromatogr. A*, 1999, 855: 455 - 486.
- [2] 王连生, 韩朔暎. 分子结构、性质与活性 [M]. 北京: 化学工业出版社, 1997.
- [3] HANNA M, DEBASIV, VERN B. Estimation of the partitioning characteristics of drugs: a comparison of a large and diverse drug series utilizing chromatographic and electrophoretic methodology[J]. *Anal Chem*, 1998, 70: 2092 - 2099.
- [4] TAN L C, CARR P W, ABRAHAM M H. Study of retention in reversed-phase Liquid chromatography using linear solvation energy relationships[J]. *J Chromatogr. A*, 1996, 752: 1 - 18.
- [5] VALKO K, PLASSN, BEVAN C, et al. Application of microemulsion electrokinetic chromatography to the analysis[J]. *J Chromatogr. A*, 1998, 797: 41 - 55.
- [6] POMPE M, NOVIC M. Prediction of gas-chromatographic retention indices using topological descriptors[J]. *J Chem Inf Comput Sci*, 1999, 39: 59 - 67.
- [7] LU H X, XUE C X, ZHANG R S, et al. Study of quantitative structure-mobility relationship of acids in capillary electrophoresis by heuristic method and radial basis function neural network[J]. *J Chem Inf Comput Sci*, 2004, 44: 1979 - 1986.
- [8] 申奇, 许禄, 李华. 新拓扑指数 A_m 在构效关系研究中的应用. 对烃类气相色谱的保留指数的分析 [J]. *分析化学*, 1996, 24(10): 1152 - 1155.
- [9] 施介华, 江峰, 严巍, 等. 烃类化合物在不同色谱柱上的定量结构 - 保留相关性研究 [J]. *分析化学*, 2005, 33(2): 181 - 186.
- [10] JOHN A D. 兰氏化学手册 [M]. 魏俊发等译. 北京: 科学出版社, 2003: 956 - 975.
- [11] 聂长明, 戴益民, 文松年, 等. 烷烃加和型性质的拓扑同系递变规律研究 [J]. *化学学报*, 2005, 63(15): 1449 - 1455.
- [12] WENER H. Structural determination of paraffin boiling points[J]. *J Am Chem Soc*, 1947, 69: 12 - 20.
- [13] GORDON M, SCANTLEBURY G R. Structural determination of paraffin properties[J]. *J Chem Soc Trans Faraday Soc*, 1964, 60: 604 - 621.
- [14] 王化云. 烷烃异构体的结构排序与热力学性质的变化规律 [J]. *化学学报*, 1993, 51(3): 216 - 223.
- [15] 冯长君, 王超. 用均根拓扑指数与路径数预测链烷烃的沸点和热力学性质 [J]. *有机化学*, 2003, 23(10): 1169 - 1176.
- [16] LU S S, CAO C Z, LIZL. Approach to estimation and prediction for normal boiling point (NBP), of alkanes based on a novel molecular distance - edge (MDE) vector[J]. *J Chem Inf Comput Sci*, 1998, 38: 387 - 394.
- [17] 顾蕙祥, 阎宝石. 气相色谱实用手册 [M]. 北京: 化学工业出版社, 1997: 68 - 70.
- [18] 李浩春. 分析化学手册, 气相色谱分析 [M]. 北京: 化学工业出版社, 1999: 357 - 482.
- [19] HEBERGER K, GORGENYIM. Principal component analysis of Kovás indices for carbonyl compounds in capillary gas chromatography[J]. *J Chromatogr. A*, 1999, 845: 21 - 31.
- [20] 许禄. 化学计量学方法 [M]. 北京: 科学出版社, 2004: 134 - 156.
- [21] LIVINGSTONE D. Data analysis for chemists[M]. Oxford: Oxford University Press, 1995: 134 - 135.
- [22] 沐来龙, 冯长君. 不饱和链烃沸点的拓扑研究 [J]. *有机化学*, 2004, 24(2): 173 - 183.
- [23] REN B Y. Atomic-level-based AI topological descriptors for structure-property correlations[J]. *J Chem Inf Comput Sci*, 2003, 43: 1121 - 1131.
- [24] NEEDHAM D E, WEI I C, SEYBOLD P G. Molecular modeling of physical properties of the alkanes[J]. *J Am Chem Soc*, 1988, 110(13): 4186 - 4194.
- [25] 卢佩章, 戴朝政, 张祥民. 色谱基础理论 [M]. 北京: 化学工业出版社, 1998: 206 - 207.