

启发式方法研究有机化合物在全二维气相色谱 中的定量结构-保留关系

纪永升^{1,2}, 夏彬彬¹, 栾 锋¹, 张晓昀¹, 张海霞^{1*}

(1. 兰州大学化学化工学院, 甘肃 兰州 730000; 2. 河南中医学院药学院, 河南 郑州 450000)

摘要 :采用半经验量子化学 PM3 的方法计算出 130 个有机化合物的描述符,用启发式方法分别对化合物在全二维气相色谱的 3 支色谱柱上的保留值建立了相应的定量结构-保留相关模型,并对模型进行了检验。所建模型呈现较好的线性,相关系数的平方(R^2)均大于 0.88,标准偏差(S)均小于 0.105,留一法交互检验的相关系数的平方(R_{cv}^2)与所建模型的 R^2 相当,说明模型具有良好的稳定性。化合物的预测结果显示所建模型有较准确的预测能力。

关键词 :全二维气相色谱;启发式方法;定量结构-保留相关;有机化合物

中图分类号 :O658 文献标识码 :A 文章编号 :1000-8713(2010)09-0826-07

Quantitative structure-retention relationship model of organic compounds on comprehensive two-dimensional gas chromatography by heuristic method

Ji Yongsheng^{1,2}, Xia Binbin¹, Luan Feng¹, Zhang Xiaoyun¹, Zhang Haixia^{1*}

(1. College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, China;
2. College of Pharmacy, Henan University of Traditional Chinese Medicine, Zhengzhou 450000, China)

Abstract :The descriptors of organic compounds were calculated by means of semi-empirical PM3 method and the quantitative structure-retention relationship model of organic compounds on three columns of comprehensive two-dimensional gas chromatography were set by heuristic method. The good stability and predictive power of the multi-linear model were tested and the results showed the model had good linear relationships with all squared correlation coefficients more than 0.88 and each standard deviation less than 0.105. The leave-one-out cross-validated squared correlation coefficients were corresponding with the squared correlation coefficients from the model.

Key words : comprehensive two-dimensional gas chromatography (GC × GC); heuristic method; quantitative structure-retention relationship; organic compounds

色谱作为一种分离技术在理论和技术方面已经相当成熟。但是当样品比较复杂时,传统的一维色谱不能满足分析要求,故而产生了多维色谱技术^[1]。全二维色谱(GC × GC)是 20 世纪 90 年代出现的新方法。Phillips 等^[2]于 1991 年在 Jorgenson 等^[3]的二维色谱思路实现了 GC × GC 的应用,并在 1999 年实现了商品化。该技术具有分辨率高、峰容量大、灵敏度高、分析时间短等特点^[1],特别适宜于复杂体系样品的分析。

GC × GC 近年来主要应用于石油产品、环境样品等复杂体系样品的分析^[4,5]。Mao 等^[6,7]用 GC × GC 对石油、石油产品的分析均得到了较好的分离; Seeley 等^[8-10]改进 GC × GC 模式为 GC × 2GC 的模式并成功地应用于挥发性有机体系分析; Xu 等^[11-13]利用 GC × GC 与质谱联用技术分别进行烟草、白酒、人参等样品的分析,均得到了较好的实验结果; Liu 等^[14]应用 GC × GC 分析了血清中的残留农药。与其他方法相比,GC × GC 对复杂样品的分

* 通讯联系人: 张海霞, 博士, 教授. Tel: (0931)4165997, E-mail: zhanghx@lzu.edu.cn.

基金项目: 国家自然科学基金项目(No. 20775029).

收稿日期: 2010-05-19

析有明显的优势。但 GC × GC 仪器价格昂贵、操作相对复杂且谱图解析较烦琐。定量结构-保留相关 (quantitative structure-retention relationship, QSRR) 方法可以解析和预测分离结果,协助优化实验参数,降低分析成本,还可通过解释实验现象,为溶质在不同色谱条件下的保留行为提供理论依据^[15]。

QSRR 是定量结构-性质相关 (quantitative structure-property relationship, QSPR) 分析方法在色谱保留研究中的具体应用^[15]。在近几十年里, QSRR 方法受到了普遍的认可^[16]。有关 QSRR 的研究国内外已有大量的文献报道^[17-21]。QSRR 方法主要包括数据的收集、分子描述符的产生和选取、相关模型的建立和评价。目前用于分子结构描述的主要有拓扑、组成、几何、静电、量子化学等描述符^[16],每类中又包含很多个描述符。因此,该方法的关键是如何选择合适的描述符建立模型。

目前,使用 QSRR 方法预测一维色谱保留数据的报道很多^[22-24],但是由于影响 GC × GC 保留行

为的参数复杂,还未见利用 QSRR 方法预测 GC × GC 保留数据的报道。本文采用启发式方法 (heuristic method, HM) 建立了化合物在 GC × GC 各色谱柱上色谱保留相关的 QSRR 模型,取得了较好的预测结果。同时也揭示了该体系中化合物结构与色谱保留之间的关系。

1 实验部分

1.1 数据集

本研究使用的数据来自文献^[9],130 个有机化合物中包括烃、醇、酮、酯类等。表 1 列出了它们在 GC × 2GC 模式全二维气相色谱系统中 DB-624 柱上的色谱保留值。该 GC × 2GC 系统包括:一维色谱柱 DB-624 (15 m × 250 μm × 1.4 μm, 6% cyanopropylphenyl, 94% dimethylpolysiloxane);二维色谱柱 DB-Wax (5 m × 250 μm × 0.25 μm, polyethyleneglycol) 和 DB-210 (5 m × 250 μm × 0.50 μm, trifluoropropylmethyl polysiloxane)。

表 1 化合物的实验和预测保留值
Table 1 Experimental and predicted retention values of the compounds

No.	Compound	log t_R (DB-624)			No.	Compound	log t_R (DB-624)		
		Pred.	Exp.	Resid.			Pred.	Exp.	Resid.
1	methane	1.6321	1.6561	-0.0240	32	<i>n</i> -propylbenzene	2.3730	2.3899	-0.0169
2	pentane	1.9937	1.8921	0.1016	33	1,3,5-trimethylbenzene	2.3453	2.3993	-0.0540
3	hexane	2.0713	2.0253	0.0460	34	<i>tert</i> -butylbenzene	2.4232	2.4160	0.0072
4	isooctane	2.1489	2.1364	0.0125	35	isobutylbenzene	2.4453	2.4259	0.0194
5	heptane	2.1489	2.1455	0.0034	36	butylbenzene	2.4517	2.4550	-0.0033
6	octane	2.2234	2.2443	-0.0209	37	styrene	2.2828	2.3520	-0.0692
7	nonane	2.2973	2.3276	-0.0303	38	methanol	2.2828	1.8182	0.4646
8	decane	2.3697	2.3990	-0.0293	39	ethanol	1.9286	1.8982	0.0304
9	undecane	2.4430	2.4601	-0.0171	40	1-propanol	2.0437	2.0358	0.0079
10	dodecane	2.5145	2.5128	0.0017	41	2-methyl-1-propanol	2.1148	2.1179	-0.0031
11	tridecane	2.5876	2.5580	0.0296	42	1-butanol	2.1405	2.1575	-0.0170
12	isoprene	2.0033	1.9196	0.0837	43	3-methyl-1-butanol	2.2286	2.2294	-0.0008
13	cyclohexane	2.0684	2.1159	-0.0475	44	1-pentanol	2.2216	2.2562	-0.0346
14	cyclohexene	2.1142	2.1418	-0.0276	45	1-hexanol	2.2216	2.3385	-0.1169
15	methylcyclohexane	2.1811	2.1830	-0.0019	46	1-heptanol	2.3742	2.4094	-0.0352
16	1,7-octadiene	2.2574	2.2360	0.0214	47	1-octanol	2.4469	2.4695	-0.0226
17	1-octene	2.2628	2.2405	0.0223	48	2-propanol	2.0123	1.9504	0.0619
18	cyclooctene	2.2691	2.3497	-0.0806	49	2-butanol	2.0920	2.0821	0.0099
19	alpha-pinene	2.2691	2.3701	-0.1010	50	2-pentanol	2.1775	2.1898	-0.0123
20	cyclooctadiene	2.2819	2.3755	-0.0936	51	3,3-dimethyl-2-butanol	2.2265	2.2271	-0.0006
21	1-decene	2.4093	2.3962	0.0131	52	4-methyl-2-pentanol	2.2448	2.2470	-0.0022
22	beta-pinene	2.4780	2.4048	0.0732	53	2-hexanol	2.2636	2.2799	-0.0163
23	delta-3-carene	2.4929	2.4218	0.0711	54	2-heptanol	2.3409	2.3577	-0.0168
24	limonene	2.4595	2.4338	0.0257	55	2-octanol	2.4175	2.4250	-0.0075
25	benzene	2.0836	2.1326	-0.0490	56	2-undecanol	2.6361	2.5760	0.0601
26	toluene	2.1948	2.2405	-0.0457	57	2-methyl-2-propanol	2.0766	1.9823	0.0943
27	ethylbenzene	2.2843	2.3216	-0.0373	58	2-methyl-2-butanol	2.1536	2.1255	0.0281
28	<i>p</i> -xylene	2.2710	2.3255	-0.0545	59	2-methyl-2-pentanol	2.2361	2.2188	0.0173
29	<i>m</i> -xylene	2.2779	2.3276	-0.0497	60	2-methyl-2-hexanol	2.3153	2.3038	0.0115
30	<i>o</i> -xylene	2.2738	2.3495	-0.0757	61	2-methyl-2-propen-1-ol	2.1125	2.1458	-0.0333
31	isopropylbenzene	2.3590	2.3700	-0.0110	62	4-penten-1-ol	2.2285	2.2512	-0.0227

表 1 (续)
Table 1 (Continued)

No.	Compound	log t_R (DB-624)			No.	Compound	log t_R (DB-624)		
		Pred.	Exp.	Resid.			Pred.	Exp.	Resid.
63	<i>cis</i> -3-hexen-1-ol	2.3058	2.3343	-0.0285	97	methyl propanoate	2.0426	2.0952	-0.0526
64	<i>trans</i> -2-hexen-1-ol	2.3166	2.3397	-0.0231	98	isopropyl acetate	2.1365	2.1290	0.0075
65	cyclohexanol	2.2859	2.3602	-0.0743	99	<i>tert</i> -butyl acetate	2.2172	2.1703	0.0469
66	4-hydroxy-4-methyl-2-pentanone	2.3959	2.3391	0.0568	100	propyl acetate	2.1696	2.1858	-0.0162
67	acetaldehyde	1.8798	1.8035	0.0763	101	isobutyl acetate	2.2543	2.2430	0.0113
68	propanal	1.9960	1.9345	0.0615	102	ethyl butanoate	2.2532	2.2655	-0.0123
69	isobutanol	2.0898	2.0191	0.0707	103	butyl acetate	2.2652	2.2776	-0.0124
70	butanal	2.1112	2.0633	0.0479	104	methyl pentanoate	2.2629	2.2865	-0.0236
71	pentanal	2.2058	2.1824	0.0234	105	isoamyl acetate	2.3662	2.3304	0.0358
72	hexanal	2.2909	2.2783	0.0126	106	4-penten-1-yl acetate	2.3702	2.3512	0.0190
73	heptanal	2.3679	2.3595	0.0084	107	pentyl acetate	2.3679	2.3570	0.0109
74	octanal	2.4436	2.4283	0.0153	108	hexyl acetate	2.4498	2.4234	0.0264
75	nonanal	2.5152	2.4871	0.0281	109	octyl acetate	2.6138	2.5311	0.0827
76	decanal	2.5879	2.5377	0.0502	110	diethyl ether	1.9870	1.9112	0.0758
77	undecanal	2.6580	2.5816	0.0764	111	1,2-dimethoxyethane	2.0956	2.1225	-0.0269
78	crotonaldehyde	2.0913	2.1477	-0.0564	112	methylal	1.9901	1.9405	0.0496
79	<i>trans</i> -2-hexenal	2.2936	2.3351	-0.0415	113	furan	1.9751	1.9232	0.0519
80	propanone	1.9871	1.9425	0.0446	114	benzaldehyde	2.3517	2.4191	-0.0674
81	butanone	2.0789	2.0745	0.0044	115	acetophenone	2.4160	2.4844	-0.0684
82	2-pentanone	2.1681	2.1767	-0.0086	116	methyl benzoate	2.4281	2.4904	-0.0623
83	4-methyl-2-pentanone	2.2605	2.2271	0.0334	117	phenylacetone	2.5159	2.5192	-0.0033
84	2-hexanone	2.2548	2.2732	-0.0184	118	benzyl acetate	2.5342	2.5235	0.0107
85	5-methyl-2-hexanone	2.3180	2.3304	-0.0124	119	ethyl benzoate	2.5145	2.5281	-0.0136
86	3-heptanone	2.3202	2.3481	-0.0279	120	cinnamaldehyde	2.5402	2.5911	-0.0509
87	2-heptanone	2.3402	2.3535	-0.0133	121	acetonitrile	1.9801	1.9647	0.0154
88	2-octanone	2.4152	2.4226	-0.0074	122	butyronitrile	2.1376	2.1833	-0.0457
89	2-nonanone	2.4890	2.4816	0.0074	123	methylene chloride	1.9574	1.9818	-0.0244
90	5-hexen-2-one	2.2510	2.2639	-0.0129	124	2-chlorobutane	2.0895	2.0667	0.0228
91	cyclopentanone	2.2094	2.2887	-0.0793	125	chloroform	2.0563	2.0959	-0.0396
92	cyclohexanone	2.2815	2.3740	-0.0925	126	1-chlorobutane	2.1119	2.1156	-0.0037
93	ethyl formate	1.9368	1.9523	-0.0155	127	1,1,1-trichloroethane	2.1606	2.1216	0.0390
94	methyl acetate	1.9207	1.9708	-0.0501	128	trichloroethylene	2.1876	2.1717	0.0159
95	vinyl acetate	2.0154	2.0366	-0.0212	129	1,2-dichloropropane	2.1770	2.1833	-0.0063
96	ethyl acetate	2.0436	2.0752	-0.0316	130	chloropentane	2.1948	2.2251	-0.0303

t_R : retention time ; Pred. : predicted value ; Exp. : experimental value ; Resid. : residual value.

1.2 分子描述符的产生

所有的分子结构均用 Hyperchem 软件画出。采用半经验的量子化学方法 PM3 优化分子结构,优化在严格的 Hartree-Fock 水平上进行,没有考虑组态相互作用,采用 Polak-Ribiere 算法直至均方根差(RMS)梯度达到 4.187 J/mol(0.001 kcal/mol)。再将优化后的分子结构文件输入到 CODESSA 软件中,分别计算出分子的拓扑、组成、几何、电子、量子化学等类别描述符^[22]。本文中对每个化合物共计算出 459 个描述符。

1.3 启发式多元线性回归法

如前所述,描述符的选择在模型建立过程中非常重要,它直接决定了模型的性能。CODESSA 软件中的 HM 是常用变量选择方法。HM 的基本过程

包括参数预选、符合条件的描述符进行相关分析和建立线性模型。该方法首先对分子描述符进行线性控制,如任意两个相关系数大于 0.8 的描述符不会同时包含在同一个模型中,并采用启发式算法对参数进行快速筛选建立模型,而不是考察所有可能的参数组合。方法采用的预处理方式是根据以下 4 条原则除掉一些描述符(1)不是每个化合物都共有的参数(2)对所有化合物来说,数值变化较小的描述符,此类描述符对分子的贡献基本一致很难体现分子结构的特异性(3)在一参数相关方程中, F 检验值小于 1.0 的参数(4) t 检验值小于某一定义值的描述符。模型的好坏由相关系数的平方(R^2)、 F 检验值以及标准偏差(S)等来检验,模型的稳定性用留一法(leave-one-out, LOO)交互检验相关系数

的平方(R_{cv}^2)来检验^[25]。用 HM 建立模型比其他方法一般要快 2~5 倍,而且所建立的模型质量也比较高^[26],因此本文使用 HM 建立多元线性回归模型。

2 结果与讨论

2.1 模型建立

将计算得到的描述符利用 HM 建立线性回归模型。为了得到合适的描述符,分别研究了不同的描述符子集,考察了描述符从 1 个增加到 10 个对模型性能的影响。当增加一个描述符而模型的性能没有显著改进时,说明描述符的个数已达到最优。为了避免模型的过参数化,我们选择 R^2 的增加值小于 0.02 作为标准。当 R^2 值随描述符个数增加的变化值小于 0.02,表明增加的描述符不能明显地改进模

型的性能,此时所选的描述符是最佳的。对于色谱柱 DB-624、DB-Wax 和 DB-210 所对应的模型,最终分别选择了 6、5、8 个描述符。我们对化合物在全二维气相色谱的 3 支色谱柱上的保留分别建立了模型,如表 2 所示,它们的 R^2 值均大于 0.88。模型中包含了拓扑、组成、几何、静电等类别的描述符,表 3 列出了各描述符及其 t 检验值。所建模型的性能与描述符在子集里的排序无关,描述符与保留值的相关性的好坏是通过每个描述符的 t 检验的绝对值来体现的,描述符的 t 检验的绝对值越大,与保留值的相关性就越大。用建立的模型对化合物的色谱保留进行了预测,预测结果列于表 1 中。图 1、图 2 和图 3 给出了 HM 得出的各模型的预测值与实验值的相关图。

表 2 化合物在 3 支色谱柱上的 QSRR 模型

Table 2 Quantitative structure-retention relationship models for the compounds on the three columns

Column	Correlation equation
DB-624	$\log t_R = 1.6006 + 0.00082(\text{GI}) + 5.7359(\text{HA}) - 1.1699(\text{RNO}) + 0.0096(\text{DPSA-3}) + 0.07907(^2\text{IC}) - 0.2198(^2\text{BIC})$ $N = 130, R^2 = 0.9005, F = 185.60, S = 0.0608, R_{cv}^2 = 0.8910$
DB-210	$t_R = 1.8338 + 0.0869(\text{HASA-2}) - 1.0504(\text{HDCA-2}) + 0.0874(\text{NO}) + 1.2308(\text{RNC}) + 2.0337(\text{MPC}_C)$ $N = 130, R^2 = 0.8892, F = 198.98, S = 0.0894, R_{cv}^2 = 0.8755$
DB-Wax	$t_R = 1.9407 + 2.9112(\text{PP/SD}) - 0.4097(\text{RNSB}) + 0.01424(\text{HDSA-1}) - 5.9194(\text{FNSA-2}) + 0.1833(\text{NR}) + 0.7873(\text{MV}) - 0.01461(\text{CHDS}) - 5.5469(\text{MPC})$ $N = 130, R^2 = 0.8846, F = 115.89, S = 0.1053, R_{cv}^2 = 0.8602$

N : number of compounds; R : correlation coefficient; F : Fisher test value; S : standard deviation; R_{cv} : cross-validated correlation coefficient.

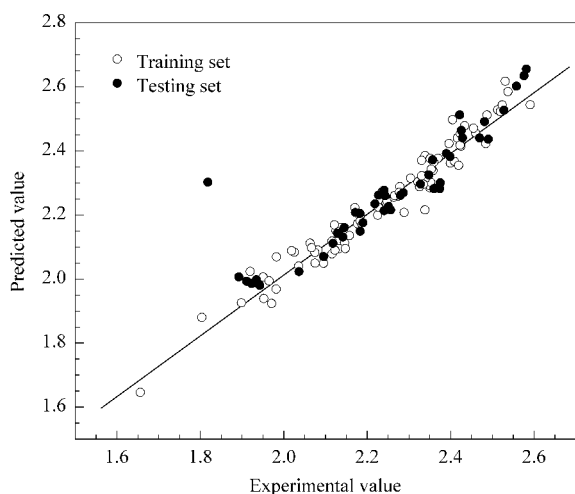
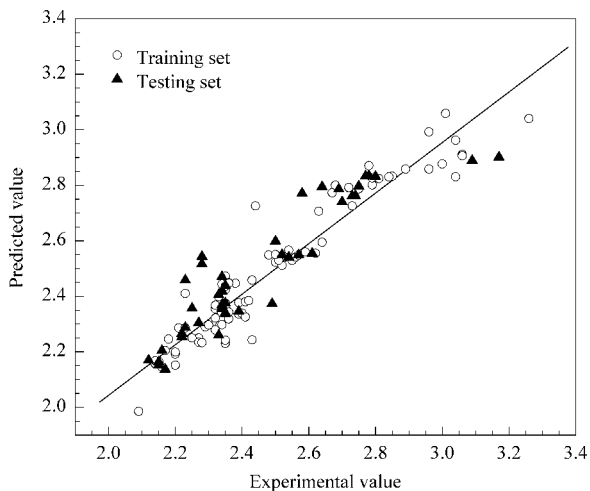
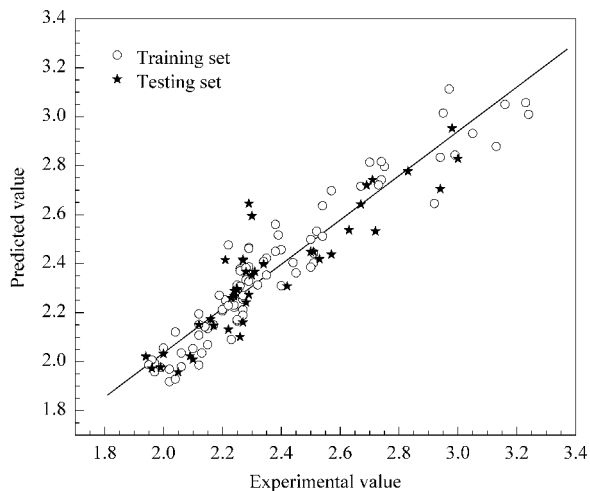
表 3 描述符及其 t -检验值

Table 3 Descriptors and their t -test values

Column	Descriptor	t -Test
DB-624	GI: gravitation index (all bonds)	16.16
	HA: HA dependent HDSA-2/TMSA [Zefirov's PC]	7.29
	RNO: relative number of O atoms	-7.91
	DPSA-3: DPSA-3 difference in CPSAs (PPSA3 - PNSA3) [Zefirov's PC]	4.26
	² IC: average information content (order 2)	3.61
	² BIC: average bonding information content (order 2)	-2.45
DB-210	HASA-2: HASA-2 [Zefirov's PC]	21.55
	HDCA-2: HA dependent HDCA-2 [Zefirov's PC]	-11.59
	NO: Number of O atoms	4.83
	RNC: relative number of C atoms	9.32
	MPC _C : max partial charge for a C atom [Zefirov's PC]	5.57
DB-Wax	PP/SD: polarity parameter/square distance	11.69
	RNSB: relative number of single bonds	3.49
	HDSA-1: HA dependent HDSA-1 [Zefirov's PC]	10.95
	FNSA-2: FNSA-2 fractional PNSA (PNSA-2/TMSA) [Zefirov's PC]	-9.32
	NR: number of rings	6.29
	MV: molecular volume/XYZ Box	4.82
	CHDS: count of H-donors sites [Zefirov's PC]	-5.83
	MPC: max partial charge (Qmax)	-5.1

色谱理论认为,气相色谱保留值主要是由溶质和固定相之间的作用力决定的,主要包括定向力、诱导力、色散力和氢键作用力等^[15];不同的色谱保留行为是溶质分子结构和色谱柱固定相性质的宏观体

现。通过解析表征分子结构信息的描述符可以建立结构与保留之间的相互关系。如果描述符在模型中存在正的系数,说明该描述符的值越大,化合物的保留越大,否则相反。

图 1 DB-624 柱上 $\log t_R$ 的实验值与预测值的相关Fig. 1 Correlation between experimental $\log t_R$ and predicted $\log t_R$ for DB-624 column图 2 DB-210 柱上 t_R 的实验值与预测值的相关Fig. 2 Correlation between experimental t_R and predicted t_R for DB-210 column图 3 DB-Wax 柱上 t_R 的实验值与预测值的相关Fig. 3 Correlation between experimental t_R and predicted t_R for DB-Wax column

在一维色谱柱 DB-624 模型中包含了 6 个描述符。这 6 个描述符中包括: 1 个组成参数——相对氧原子数目(RNO)。RNO 与分子的极化作用相关,且在模型中的系数为负值。这与色谱理论一致,因为氧原子电负性较大,氧原子数目增加能使分子极性增强,使化合物在弱极性固定相 DB-624 上的保留减小。2 个静电参数——加权氢键给体原子的溶剂可极化表面积(HA)和偏电荷表面积差量(DPSA-3)。HA 是由原子偏电荷和分子溶剂化表面积等信息描述分子氢键作用力;DPSA-3 表示分子正偏电荷极化表面积与负偏电荷极化表面积之间的关系,与分子的极化作用密切相关;HA 和 DPSA-3 都属于静电参数中分子偏电荷与表面积描述符类型(charged partial surface area descriptors, CP-SAs),在模型中的系数都是正值,说明化合物电荷分布可以影响化合物在弱极性固定相上的保留。2 个拓扑参数——二阶键的平均信息指数(2 BIC)和二阶平均信息指数(2 IC);它们描述了分子的形状、分支度和分子组成等信息,与分子的色散作用关系密切^[27]。1 个几何参数——所有键的重力指数(GI)。GI 与相对原子质量的大小和原子的分子空间分布有关,可以定量描述分子的色散和疏水作用对色谱保留的影响^[17],其在模型中的系数为正;从表 3 中可以看出 GI 的 t 检验值最大,说明化合物在 DB-624 柱上的色谱保留与化合物分子的色散力最相关,这与文献[9]的解释完全一致。

在二维色谱柱 DB-210 的模型中有 5 个描述符。包括 3 个静电参数——加权氢键受体原子的可溶剂化表面积(HASA-2)、氢键给体原子的溶剂化表面积(HDCA-2)和碳原子的最大偏电荷(MPC_C)。其中 HASA-2 和 HDCA-2 都属 CP-SAs,都可以表征溶质分子与固定相电荷作用对色谱保留的影响。2 个组成描述参数——氧原子数目(NO)和相对碳原子数(RNC)。RNC 表示碳原子数与原子总数的比值,反映了分子的组成和结构特征,可以表征分子与固定相的位阻作用对色谱保留的影响。在这个模型中,除 HDCA-2 描述符具有负的系数外,其他都为正的系数,并且 HASA-2 描述符与保留值关系最大,这可以从固定相的结构得到解释。固定相上存在极化作用强并可以构建氢键的受体 F 原子,显然在气相色谱中只有溶质可以与之形成氢键,而与载气无关。所以在这个固定相上化合物的保留将与氢键能力最相关。另外一个二维色谱柱 DB-Wax 的模型中有 5 个静电描述符^[25]。包括极化参数(PP/SD)、加权负偏电荷溶剂可极化面积(FNSA-2)、氢键给体原子的

溶剂可极化面积 (HDSA-1)、氢键给体位点数 (CHDS) 和最大偏电荷 (MPC)。它们分别描述分子极化、氢键作用、电荷分布信息,可以定量表征分子的极化、氢键作用。2 个组成参数——相对单键数 (RNSB) 和分子中环的数目 (NR)。RNSB 值的多少会影响分子电子云的密度,也是影响极化作用的重要因素^[28]。NR 反映了分子的组成和立体结构信息,对分子的位阻和疏水作用影响较大。一个几何参数——分子体积 (MV)。MV 表征分子的三维立体空间信息,其对分子的色散力和空间位阻作用影响较大,结构相似的分子的 MV 大,色散力也较大^[28]。从表 3 的 t 检验值不难看出,对 DB-Wax 柱,氢键和极化作用是影响保留的两个最重要的参数。在文献 [9] 中,Seeley 等通过考察实验现象认为,影响化合物在 DB-210 色谱柱上的保留主要是极化作用;而对于 DB-Wax 色谱柱,则认为是氢键作用为主。通过计算,我们可以发现,相对于极化作用,氢键作用更能影响化合物在 DB-210 色谱柱上的保留,说明 F 原子更倾向于形成氢键而非极化作用。在 DB-Wax 色谱柱上,氢键与极化作用对保留的影响都较大。

以上模型中所选的描述符有明确的意义,它们可以反映分子结构特征与色谱保留的关系,且与色谱理论和文献 [9] 的有关保留行为的解释基本吻合。与一维色谱相比,全二维色谱的定量结构-保留关系的研究,能更深入透彻地理解化合物的色谱行为和保留机理。另外,全二维色谱的 QSRR 模型可以实现化合物的二维数据的模拟预测。图 4 为对几种烃类化合物的二维色谱模拟,这对二维色谱大量复杂的数据处理分析会有很大的帮助。

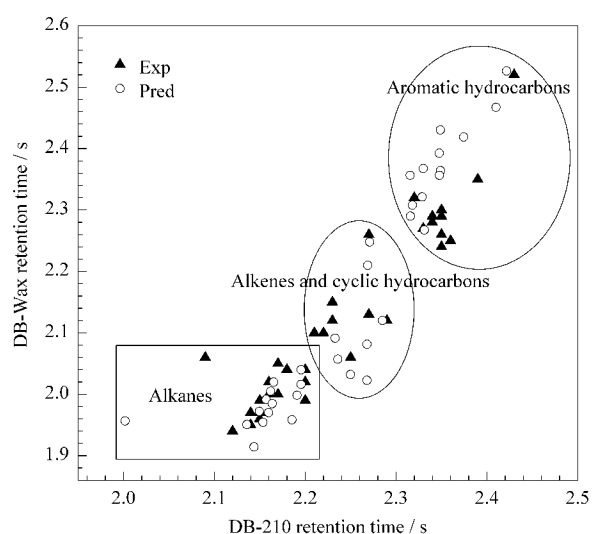


图 4 烷烃在 DB-Wax 和 DB-210 上的二维保留模拟
Fig. 4 DB-Wax retention times for hydrocarbons of VOCs plotted as a function of DB-210 retention times

2.2 模型的检验

如表 2 所示,模型的留一法检验的相关系数平方 R_{cv}^2 均大于 0.86,且与 R^2 接近,说明模型较稳定。为了进一步检验模型的稳定性和所选参数的合理性,我们把样本随机分为 A、B、C 3 组。A 组化合物编号为 1, 4, 7, 10, ..., 130; B 组化合物编号为 2, 5, 8, 11, ..., 128; C 组化合物编号为 3, 6, 9, 12, ..., 129。从中任意选两组作训练集,余下一组作测试集。分别用表 2 中 3 个模型所选的参数作为自变量,采用多元线性回归分析进行交互检验,结果如表 4 所示,所有模型交互检验相关系数平方 R^2 的平均值均大于 0.88,与所建模型的相当。表 4 的结果说明所建立的模型均有较好的稳定性和预测能力,同时也说明所选的参数合理。

表 4 交互检验结果
Table 4 Results of cross-validated test

Column	Training set	R^2	S	Testing set	R^2	S
DB-624	B + C	0.8726	0.0612	A	0.9494	0.0460
	A + C	0.9493	0.0435	B	0.8103	0.0785
	A + B	0.8866	0.0685	C	0.9344	0.0421
	Average	0.9028	0.0616		0.8980	0.0556
DB-210	B + C	0.8915	0.0900	A	0.8847	0.0878
	A + C	0.9105	0.0830	B	0.8449	0.0913
	A + B	0.8707	0.0943	C	0.9240	0.0727
	Average	0.8909	0.0888		0.8845	0.0839
DB-Wax	B + C	0.8885	0.102	A	0.8716	0.0983
	A + C	0.9047	0.101	B	0.8155	0.1104
	A + B	0.8701	0.111	C	0.9074	0.0986
	Average	0.8878	0.105		0.8648	0.1024

A-group: compounds No. 1, 4, 7, 10, ..., 130; B-group: compounds No. 2, 5, 8, 11, ..., 128; C-group: compounds No. 3, 6, 9, 12, ..., 129.

3 结论

用 HM 建立了包括烷烃、烯烃、酮、酯等多种类型的 130 个有机化合物的 QSRR 模型,得到了较好的结果。所建线性模型显示了好的预测能力和泛化能力,并可以直观地看出与化合物的色谱保留有关的结构特征。该方法不仅可以帮助分析复杂样品的色谱保留值,而且可以预测化合物在全二维色谱的保留行为,为设计实验、寻找合适的色谱柱提供帮助。

参考文献:

- [1] Xu G W , Ye F , Kong H W , et al. Chinese Journal of Chromatography (许国旺,叶芬,孔宏伟,等. 色谱), 2001 , 19 (2): 132
- [2] Liu Z , Phillips J B. J Chromatogr Sci , 1991 , 29 : 227
- [3] Bushey M M , Jorgenson J W , Anal Chem , 1990 , 62 : 161
- [4] Mühlen C V , Zini C A , Caramão E B , et al. J Chromatogr A , 2008 , 1200 : 34
- [5] Eiceman G , Gardea-Torresdey J , Overton E D , et al. Anal Chem , 2004 , 76 : 3387
- [6] Mao D , Van De Weghe D , Diels L , et al. J Chromatogr A , 2008 , 1179 : 33
- [7] Dutriez T , Courtiade M , Thiébaud D , et al. Fuel , 2010 , 89 : 2338
- [8] Seeley J V , Kramp F , Hicks C J. Anal Chem , 2000 , 72 : 4346
- [9] Seeley J V , Kramp F , Sharpe K S. J Sep Sci , 2001 , 24 : 444
- [10] Seeley J V , Kramp F , Sharpe K S , et al. J Sep Sci , 2002 , 25 : 53
- [11] Lu X , Zhao M Y , Kong H W , et al. J Sep Sci , 2004 , 27 : 101
- [12] Ryan D , Watkins P , Smith J , et al. J Sep Sci , 2005 , 28 : 1075
- [13] Shellie R A , Marriott P J , Huie C W. J Sep Sci , 2003 , 26 : 1185
- [14] Liu Z , Sirimanne S R , Patterson D G Jr , et al. Anal Chem , 1994 , 66 : 3086
- [15] Shi J H , Jiang F , Yan W , et al. Chinese Journal of Analytical Chemistry (施介华,江峰,严巍,等. 分析化学), 2005 , 33(2) : 181
- [16] Luan F , Xue C X , Zhang R S , et al. Anal Chim Acta , 2005 , 537 : 101
- [17] Katritzky A R , Chen K. Anal Chem , 2000 , 72 : 101
- [18] Xu H , Zou J , Jiang Y , et al. J Chromatogr A , 2008 , 1198 : 202
- [19] Put R , Daszykowski M , Baczek T , et al. J Proteome Res , 2006 , 5 : 1618
- [20] Liu H Y , Wang Z Y , Liu S S , et al. Chinese Journal of Chromatography (刘红艳,王遵尧,刘树深,等. 色谱), 2005 , 23(4) : 336
- [21] Zhang Q , Zhao H X , Chen J P , et al. Chinese Journal of Chromatography (张青,赵洪霞,陈吉平,等. 色谱), 2005 , 23(5) : 441
- [22] Zhou P , Tian F , Lv F , et al. J Chromatogr A , 2009 , 1216 : 3107
- [23] Wang C , Skibic M J , Higgs R E , et al. J Chromatogr A , 2009 , 1216 : 5030
- [24] Qin L , Liu S , Liu H , et al. J Chromatogr A , 2009 , 1216 : 5302
- [25] Katritzky A R , Lobanov V S , Karelson M. CODESSA : Reference Manual. Gainesville , FL , USA : University of Florida , 1994
- [26] Katritzky A R , Petrukhin R , Jain R , et al. J Chem Inf Comput Sci , 2001 , 41 : 1521
- [27] Luan F , Ma W P , Zhang X Y , et al. Chemosphere , 2006 , 63 : 1142
- [28] Xu H Y , Yu Q S , Zou J W , et al. Chinese J Struct Chem , 2006 , 25(7) : 811