

多环芳香硫化化合物的定量结构-气相色谱保留指数关系

李正华, 程凡圣, 夏之宁*

(重庆大学化学化工学院, 重庆 400030)

摘要 :应用分子电性距离矢量(MEDV)对 114 个多环芳香硫化化合物(PASHs)进行结构表征,通过多元线性回归建立了 PASHs 的气相色谱保留指数与 MEDV 参数之间的定量结构-保留值关系模型;同时采用逐步回归分析进行变量筛选,继而以留一法交互检验对所得优化模型进行预测能力评价,所建立的模型的相关系数为 0.994 7,交互检验相关系数为 0.994 0,表明该优化模型具有良好的稳定性和预测能力。此外,通过将样本集按 2:1 分成校准集和测试集预测,统计分析结果显示所建的模型具有良好的相关性和稳定性。本文所建的定量结构-保留值关系(QS-RR)模型为预测 PASHs 的气相色谱保留指数提供了一个便捷有效的新方法。

关键词 :气相色谱保留指数;定量结构-保留指数关系;分子电性距离矢量;多环芳香硫化化合物

中图分类号 :O658 文献标识码 :A 文章编号 :1000-8713(2011)01-0063-07

Quantitative structure-gas chromatographic retention relationship of polycyclic aromatic sulfur heterocycles using molecular electronegativity-distance vector

LI Zhenghua, CHENG Fansheng, XIA Zhining*

(College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400030, China)

Abstract : The chemical structures of 114 polycyclic aromatic sulfur heterocycles (PASHs) have been studied by molecular electronegativity-distance vector (MEDV). The linear relationships between gas chromatographic retention index and the MEDV have been established by a multiple linear regression (MLR) model. The results of variable selection by stepwise multiple regression (SMR) and the powerful predictive abilities of the optimization model appraised by leave-one-out cross-validation showed that the optimization model with the correlation coefficient (R) of 0.994 7 and the cross-validated correlation coefficient (R_{CV}) of 0.994 0 possessed the best statistical quality. Furthermore, when the 114 PASHs compounds were divided into calibration and test sets in the ratio of 2:1, the statistical analysis showed our models possesses almost equal statistical quality, the very similar regression coefficients and the good robustness. The quantitative structure-retention relationship (QSRR) model established may provide a convenient and powerful method for predicting the gas chromatographic retention of PASHs.

Key words : gas chromatographic retention index; quantitative structure-retention relationship (QSRR); molecular electronegativity-distance vector; polycyclic aromatic sulfur heterocycles

多环芳香硫化化合物(polycyclic aromatic sulfur heterocycles, PASHs)是原油中有机硫的主要存在形态。它的存在会对石油产品质量产生不利影响,并在燃烧过程中造成大气污染^[1]。有研究表明,一些 PASHs 有着比多环芳烃化合物更大的致突变和致癌毒性^[2]。气相色谱与质谱或原子发射检测器

联用技术已经广泛应用到石油和环境样品中的 PASHs 检测研究中^[3-6]。然而,由于 PASHs 存在大量的异构体以及缺少 PASHs 标准品,使得通过实验测定所有可能存在的 PASHs 的气相色谱保留指数十分困难。研究表明,通过定量结构-保留值关系(quantitative structure-retention relationship, QS-

* 通讯联系人:夏之宁,博士,教授。E-mail:chem_lab_cqu@yahoo.com.cn.

基金项目:国家自然科学基金(No. 20775096)和科技部国际合作项目(No. 2010DFA32680).

收稿日期:2010-09-21

RR)模型预测气相色谱保留值已经成为一种可行的方法^[7-11]。

本研究采用分子电性距离矢量(molecular electronegativity-distance vector, MEDV)对 114 个 PASHs 进行结构表征,通过多元线性回归建立 PASHs 结构与其气相色谱保留指数的 QSRR 模型,同时采用逐步回归进行变量筛选,继而以留一法交互校验(LOO-CV)进行模型稳定性和预测能力的检验。为进一步检验模型的稳定性和预测能力,将 114 种 PASHs 以 2:1 的比例分为校准集和测试集,统计分析结果显示所建模型具有良好的稳定性和预测能力。本研究中所建立的 QSRR 模型为预测多环芳香硫化物的气相色谱保留指数提供了一个便捷有效的新方法。

1 分子电性距离矢量计算

MEDV 是一种矢量结构描述子,它能很好地表达有机分子中各非氢原子的不同结构特征,以及模拟各非氢原子之间的相互作用。由于分子性质取决于分子中不同环境原子之间的相互作用,所以 MEDV 描述子具有良好的性质相关性^[12],其已经应用到多溴联苯醚^[13]、多氯萘^[14]、稠环芳烃^[15]等的

QSRR 研究中,并取得良好的效果。

MEDV 计算过程中需将有机分子中的原子进行类型和属性的划分,并定义相对电性和相对键长^[12]。各描述子的计算中借鉴库仑定理的形式,将有机分子中非氢原子按其所连接的非氢原子数目分为 4 类,各类原子发生相互作用以以下几种方式组合:

$$M_{kl} = \sum_{i \in k, j \in l} \frac{q_i q_j}{d_{ij}^2} \quad (k=1, 2, 3, 4; k \leq l \leq 4)$$

式中 M_{kl} 表示第 k 类原子和第 l 类原子的作用项。 k 或 l 为原子类型,原子 i 和 j 分别属于第 k 类原子和第 l 类原子; q_i 和 q_j 为原子 i 和 j 的相对电性; d_{ij} 表示原子 i 和 j 之间的距离(以相对键长表示),是原子通过一个或多个化学键连接到其他原子的所有路径中各个相对键长加和的最小值。这样就得 10 个变量: M_{11} 、 M_{12} 、 M_{13} 、 M_{14} 、 M_{22} 、 M_{23} 、 M_{24} 、 M_{33} 、 M_{34} 、 M_{44} ,即为 MEDV 描述子。统称为 V_1 、 V_2 、 V_3 、 V_4 、 V_5 、 V_6 、 V_7 、 V_8 、 V_9 、 V_{10} ,具体计算方法参见文献[12-15]。由于样品集中的 PASHs 分子不含第 4 类原子的非氢原子,相应的 MEDV 描述子 M_{14} 、 M_{24} 、 M_{34} 、 M_{44} 实际为零,因此对于样品集中的 PASHs 分子, MEDV 只有 6 个非零元素,其分别对应为 V_1 、 V_2 、 V_3 、 V_5 、 V_6 、 V_8 ,具体参数列于表 1。

表 1 多环芳香硫化物的 MEDV 元素和气相色谱保留指数的实验值与预测值

Table 1 MEDV descriptors together with the experimental and predicted retention index (RI) values for PASHs

No.	Compound	V_1	V_2	V_3	V_5	V_6	V_8	$RI_{obs}^{1)}$	$RI_{pred-1}^{2)}$	$RI_{pred-2}^{3)}$
1	benzo[b]thiophene	0.0000	0.0000	0.0000	8.3229	7.0709	1.2277	201.57	206.16	204.31
2	7-methylbenzo[b]thiophene	0.0000	0.7192	1.4034	6.1631	7.6960	2.7624	219.16	222.88	222.86
3	2-methylbenzo[b]thiophene	0.0000	0.7217	1.2235	5.8970	8.9871	1.7375	220.76	225.82	225.67
4	5-methylbenzo[b]thiophene	0.0000	0.8431	1.1998	5.2779	9.6726	1.6711	222.09	226.75	227.01
5	6-methylbenzo[b]thiophene	0.0000	0.8323	1.1998	5.3030	9.6475	1.6711	222.11	226.65	226.88
6	3-methylbenzo[b]thiophene	0.0000	0.7074	1.4034	6.1889	7.6703	2.7624	217.54	222.77	222.72
7	4-methylbenzo[b]thiophene	0.0000	0.7420	1.4034	6.0894	7.7698	2.7624	226.79	223.08	223.13
8	5-ethylbenzo[b]thiophene	0.0000	1.4133	0.3642	6.1210	10.8724	1.6711	243.89	242.28	243.96
9	3,5-dimethylbenzo[b]thiophene	0.0451	1.4050	2.7487	3.2803	9.9990	3.3422	236.18	236.33	237.97
10	naphtho[1,2-b]thiophene	0.0000	0.0000	0.0000	10.4858	12.7687	4.4335	298.38	298.32	296.81
11	dibenzothiophene	0.0000	0.0000	0.0000	10.9121	12.1208	4.4999	288.41	296.74	295.15
12	naphtho[2,1-b]thiophene	0.0000	0.0000	0.0000	10.4414	12.8674	4.4335	314.82	298.82	297.32
13	naphtho[2,3-b]thiophene	0.0000	0.0000	0.0000	9.7873	14.4277	3.3422	310.07	302.40	300.76
14	5-methylnaphtho[2,1-b]thiophene	0.0000	0.7486	1.6032	8.2997	13.0312	6.4115	322.27	316.16	316.58
15	2-methylnaphtho[2,1-b]thiophene	0.0000	0.7391	1.3433	8.0196	14.5663	5.1564	308.34	319.12	319.36
16	4-methyldibenzothiophene	0.0000	0.7356	1.5350	8.7625	12.4968	6.2735	304.21	314.11	314.39
17	8-methylnaphtho[1,2-b]thiophene	0.0000	0.8646	1.3197	7.4378	15.1601	5.0900	316.76	319.59	320.24
18	2-methyldibenzothiophene	0.0000	0.8604	1.3197	7.8712	14.5052	5.1564	315.70	317.97	318.52
19	3-methyldibenzothiophene	0.0000	0.8528	1.2865	7.8768	14.5743	5.0817	311.54	317.85	318.38
20	4-methylnaphtho[1,2-b]thiophene	0.0000	0.7419	1.6032	8.3566	12.9198	6.4115	310.30	315.59	315.98
21	4-methylnaphtho[2,1-b]thiophene	0.0000	0.7191	1.6032	8.3860	12.9448	6.4115	320.10	315.89	316.23
22	6-methylnaphtho[1,2-b]thiophene	0.0000	0.7809	1.5232	8.2116	13.2950	6.1813	317.74	316.08	316.57
23	1-methyldibenzothiophene	0.0000	0.7419	1.6032	8.7830	12.2720	6.4779	317.12	314.01	314.32
24	8-methylnaphtho[2,1-b]thiophene	0.0000	0.8705	1.3197	7.3819	15.2705	5.0900	314.46	320.15	320.82

表 1 (续)
Table 1 (Continued)

No.	Compound	V_1	V_2	V_3	V_5	V_6	V_8	$RI_{obs}^{1)}$	$RI_{pred-1}^{2)}$	$RI_{pred-2}^{3)}$
25	7-methylnaphtho[2 ,1- <i>b</i>]thiophene	0.0000	0.8797	1.2798	7.3507	15.3890	5.0027	316.23	320.17	320.86
26	6-methylnaphtho[2 ,1- <i>b</i>]thiophene	0.0000	0.7869	1.5232	8.1557	13.4053	6.1813	323.00	316.64	317.15
27	1-methylnaphtho[2 ,1- <i>b</i>]thiophene	0.0000	0.7073	1.6032	8.4118	12.9190	6.4115	321.26	315.78	316.09
28	9-methylnaphtho[2 ,1- <i>b</i>]thiophene	0.0000	0.7593	1.6032	8.2746	13.0562	6.4115	327.38	316.25	316.70
29	3-ethyl dibenzothiophene	0.0000	1.4273	0.4235	8.7296	15.8608	5.0817	346.20	334.00	335.98
30	4- β -dimethyl dibenzothiophene	0.0275	1.3920	3.1491	6.6747	12.7493	8.1088	324.46	327.27	329.19
31	2- β -dimethyl dibenzothiophene	0.0236	1.5302	2.9204	5.7708	14.7830	6.9791	335.34	331.74	333.97
32	2-ethyl dibenzothiophene	0.0000	1.4335	0.4411	8.7316	15.8249	5.1564	342.56	334.30	336.31
33	3- β -dimethyl dibenzothiophene	0.0208	1.5315	2.8784	5.7684	14.8679	6.8965	328.95	332.06	334.29
34	2- δ -dimethyl dibenzothiophene	0.0236	1.6550	2.7051	4.8794	16.7914	5.8620	340.49	335.60	338.11
35	3- γ -dimethyl dibenzothiophene	0.0163	1.6627	2.6159	4.8710	16.9690	5.6929	331.51	336.49	339.02
36	3- δ -dimethyl dibenzothiophene	0.0182	1.6646	2.6547	4.8700	16.8905	5.7723	333.17	336.32	338.85
37	1- γ -dimethyl dibenzothiophene	0.0236	1.5288	2.9555	5.7968	14.6272	7.1088	335.20	331.52	333.76
38	phenanthro[4 ,5- <i>bcd</i>]thiophene	0.0000	0.0000	0.0000	9.7215	16.5326	8.5923	332.60	345.39	345.19
39	phenaleno[6 ,7- <i>bc</i>]thiophene	0.0000	0.0000	0.0000	9.2750	17.2173	8.5260	345.20	347.12	347.01
40	benzo[<i>b</i>]naphtho[2 ,1- <i>d</i>]thiophene	0.0000	0.0000	0.0000	13.1458	17.8289	8.0830	384.15	391.40	390.24
41	benzo[<i>b</i>]naphtho[1 ,2- <i>d</i>]thiophene	0.0000	0.0000	0.0000	13.1882	17.7004	8.3621	390.40	391.82	390.73
42	phenanthro[9 ,10- <i>b</i>]thiophene	0.0000	0.0000	0.0000	13.4908	16.6001	9.6173	387.49	389.78	388.96
43	phenanthro[4 ,3- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.7609	18.3679	8.2958	379.63	393.58	392.57
44	anthra[1 ,2- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.0062	20.2054	6.8869	407.03	397.42	396.20
45	benzo[<i>b</i>]naphtho[2 ,3- <i>d</i>]thiophene	0.0000	0.0000	0.0000	12.4423	19.5001	6.9659	401.03	395.43	394.14
46	phenanthro[1 ,2- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.6592	18.6578	7.9782	395.38	393.95	392.88
47	phenanthro[3 ,4- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.7079	18.5033	8.2958	398.66	394.35	393.35
48	anthra[2 ,1- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.9604	20.3220	6.8869	409.90	398.08	396.87
49	phenanthro[2 ,1- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.6134	18.7744	7.9782	406.22	394.61	393.55
50	phenanthro[3 ,2- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.9607	20.3168	6.8869	404.55	398.03	396.83
51	phenanthro[2 ,3- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.9593	20.3346	6.8869	407.89	398.19	396.99
52	1-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.7767	1.7230	11.0255	17.6720	10.5533	420.94	409.89	410.78
53	11-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.7593	1.7230	11.0632	17.6343	10.5533	424.88	409.73	410.57
54	10-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.7556	1.6030	10.9771	18.1211	9.9595	397.72	409.33	410.08
55	3-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.8940	1.3255	10.0414	20.3006	8.7157	409.09	413.15	414.22
56	anthra[2 ,3- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.2891	21.9279	5.7083	409.54	401.54	400.17
57	2-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.8843	1.3810	10.0775	20.1504	8.8298	405.12	413.20	414.24
58	8-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.8797	1.3769	10.0847	20.1502	8.8227	398.56	413.14	414.17
59	9-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.8712	1.3364	10.0905	20.2318	8.7354	416.21	412.98	413.98
60	2-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.8907	1.3996	10.1122	19.9941	9.1445	409.88	413.74	414.86
61	8-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.7558	1.6149	11.0222	17.9671	10.2616	407.02	409.79	410.61
62	10-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.8806	1.3996	10.1308	19.9755	9.1445	406.98	413.65	414.74
63	6-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.7355	1.7348	11.1431	17.5287	10.5791	408.01	409.55	410.32
64	5-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.7767	1.7230	10.9831	17.8004	10.2742	403.32	409.47	410.30
65	3-methylbenzo[<i>b</i>]naphtho[1 ,2- <i>d</i>] thiophene	0.0000	0.8990	1.3370	10.0773	20.1590	9.0145	412.18	413.67	414.81
66	4-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.7554	1.5963	10.2724	19.8061	8.8298	407.42	413.34	413.95

表 1 (续)
Table 1 (Continued)

No.	Compound	V_1	V_2	V_3	V_5	V_6	V_8	$RI_{obs}^{1)}$	$RI_{pred-1}^{2)}$	$RI_{pred-2}^{3)}$
67	4-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.8007	1.5845	10.8513	18.2852	9.9211	407.56	409.69	410.57
68	6-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.7418	1.8031	11.1210	17.4323	10.5044	397.98	409.02	409.76
69	9-methylbenzo[<i>b</i>]naphtho[1 2- <i>d</i>] thiophene	0.0000	0.8720	1.3437	10.1328	20.0909	9.0271	416.06	413.44	414.51
70	7-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.7621	1.6832	11.0002	17.8706	10.1869	398.18	409.26	410.05
71	10-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8071	1.6031	10.1401	19.9287	8.8396	422.17	413.85	414.60
72	9-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8990	1.3370	9.3313	21.9587	7.6183	416.29	417.29	418.22
73	1-methylbenzo[<i>b</i>]naphtho[2 ,1- <i>d</i>] thiophene	0.0000	0.7690	1.6899	10.9887	17.8695	10.1995	407.03	409.35	410.15
74	8-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8940	1.3255	9.3379	21.9718	7.5986	417.00	417.19	418.11
75	2-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8797	1.3769	9.3812	21.8215	7.7056	418.56	417.18	418.06
76	6-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.6519	1.9383	11.1709	17.4632	10.2742	419.53	409.65	410.06
77	3-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8706	1.3322	9.3871	21.9100	7.6112	413.96	416.99	417.85
78	4-methylbenzo[<i>b</i>]naphtho[1 2- <i>d</i>] thiophene	0.0000	0.8071	1.6031	10.8860	18.1290	10.2358	414.74	410.23	411.19
79	1-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.7621	1.6832	10.2966	19.5419	9.0698	410.82	413.30	413.95
80	7-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.8007	1.5845	10.1478	19.9565	8.8040	413.86	413.73	414.46
81	3-methylphenanthro[9 ,10- <i>b</i>] thiophene	0.0000	0.7247	1.7230	11.4653	16.4345	11.8084	403.18	407.39	408.40
82	1-methylanthra[2 ,1- <i>b</i>]thiophene	0.0000	0.7275	1.6832	9.9143	20.2643	8.9908	432.18	415.64	416.28
83	10-methylphenanthro[2 ,1- <i>b</i>] thiophene	0.0000	0.7485	1.8031	10.5760	18.3904	10.3996	427.06	412.29	413.15
84	11-methylbenzo[<i>b</i>]naphtho[2 ,3- <i>d</i>] thiophene	0.0000	0.6582	2.0066	11.1913	17.2384	10.4786	418.78	409.55	409.99
85	3-methylphenanthro[2 ,1- <i>b</i>] thiophene	0.0000	0.7275	1.6832	10.5672	18.7167	10.0821	418.51	412.17	412.95
86	2-(2'-naphthyl)benzo[<i>b</i>]thiophene	0.0000	0.0000	0.0000	15.2707	22.2980	5.2422	443.62	444.52	442.07
87	benzo[2 ,3]phenanthro[4 ,5- <i>bcd</i>] thiophene	0.0000	0.0000	0.0000	12.0555	21.9226	12.7447	433.90	440.50	440.87
88	pyrenol[4 ,5- <i>b</i>]thiophene	0.0000	0.0000	0.0000	12.2927	20.9301	14.1360	442.27	439.44	440.12
89	benzo[1 ,2]phenaleno[3 ,4- <i>bc</i>] thiophene	0.0000	0.0000	0.0000	12.3656	20.7672	14.1360	432.08	438.61	439.28
90	triphenaleno[4 ,5- <i>bcd</i>]thiophene	0.0000	0.0000	0.0000	12.7594	20.2189	14.2024	430.26	437.66	438.25
91	pyrenol[1 2- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.5184	22.8498	12.6399	440.85	443.49	443.95
92	chryseno[4 ,5- <i>bcd</i>]thiophene	0.0000	0.0000	0.0000	12.0049	22.0462	12.7189	436.55	441.07	441.43
93	pyrenol[2 ,1- <i>b</i>]thiophene	0.0000	0.0000	0.0000	11.4736	22.9476	12.6399	447.02	443.97	444.45
94	benzo[4 ,5]phenaleno[1 ,9- <i>bc</i>] thiophene	0.0000	0.0000	0.0000	11.5472	22.8063	12.6399	458.77	443.36	443.82
95	benzo[4 ,5]phenaleno[9 ,1- <i>bc</i>] thiophene	0.0000	0.0000	0.0000	11.5472	22.8063	12.6399	450.66	443.36	443.82
96	benzo[<i>b</i>]phenanthro[4 ,3- <i>d</i>] thiophene	0.0000	0.0000	0.0000	15.5186	23.3612	12.5634	493.39	489.75	489.23
97	dinaphtho[2 ,1- <i>b</i> :1' 2'- <i>d</i>]thiophene	0.0000	0.0000	0.0000	15.5283	23.3049	12.5634	501.12	489.31	488.78
98	dinaphtho[1 2- <i>b</i> :2' 1'- <i>d</i>]thiophene	0.0000	0.0000	0.0000	15.4142	23.6154	11.8396	474.09	487.95	487.26
99	benzo[1 ,2]phenaleno[4 ,3- <i>bc</i>]	0.0000	0.0000	0.0000	12.3213	20.8660	14.1360	443.73	439.11	439.79

表 1 (续)
Table 1 (Continued)

No.	Compound name	V_1	V_2	V_3	V_5	V_6	V_8	$RI_{obs}^{1)}$	$RI_{pred-1}^{2)}$	$RI_{pred-2}^{3)}$
100	dinaphtho[1 2- <i>b</i> :1' 2'- <i>d</i>]thiophene	0.0000	0.0000	0.0000	15.4618	23.4819	12.1543	478.14	488.53	487.92
101	benzo[<i>b</i>]phenanthro[9 ,10- <i>d</i>] thiophene	0.0000	0.0000	0.0000	16.3085	21.4434	13.9233	484.41	485.28	484.95
102	benzo[<i>b</i>]phenanthro[3 4- <i>d</i>] thiophene	0.0000	0.0000	0.0000	15.4636	23.4995	12.1740	470.60	488.81	488.20
103	anthra[1 2- <i>b</i>]benzo[<i>d</i>]thiophene	0.0000	0.0000	0.0000	14.6982	25.3461	10.6903	483.60	492.29	491.46
104	benzo[<i>b</i>]phenanthro[2 ,1- <i>d</i>] thiophene	0.0000	0.0000	0.0000	14.7471	25.2285	11.0246	479.40	493.13	492.38
105	dinaphtho[1 2- <i>b</i> :2' 3'- <i>d</i>]thiophene	0.0000	0.0000	0.0000	14.7079	25.2887	10.7029	488.52	491.89	491.06
106	9 ,13- <i>H</i> -triphenyleno[2 3- <i>b</i>] thiophene	0.0000	0.0000	0.0000	15.0192	24.2587	12.4097	511.39	492.62	492.18
107	benzo[<i>b</i>]phenanthro[3 2- <i>d</i>] thiophene	0.0000	0.0000	0.0000	14.6541	25.4805	10.7197	488.76	493.27	492.46
108	benzo[<i>b</i>]phenanthro[1 2- <i>d</i>] thiophene	0.0000	0.0000	0.0000	15.4000	23.6808	12.1159	491.34	489.66	489.05
109	benzo[<i>b</i>]phenanthro[2 3- <i>d</i>] thiophene	0.0000	0.0000	0.0000	14.6476	25.4698	10.6645	490.44	492.86	492.03
110	triphenaleno[2 ,1- <i>b</i>]thiophene	0.0000	0.0000	0.0000	15.7663	22.4452	13.8185	500.87	488.94	488.71
111	triphenaleno[1 2- <i>b</i>]thiophene	0.0000	0.0000	0.0000	15.8208	22.2920	13.8185	486.04	488.02	487.77
112	dinaphtho[2 3- <i>b</i> :2' 3'- <i>d</i>]thiophene	0.0000	0.0000	0.0000	14.0034	26.9596	9.5787	488.87	495.88	494.91
113	triphenaleno[2 3- <i>b</i>]thiophene	0.0000	0.0000	0.0000	15.0192	24.2587	12.4097	490.89	492.62	492.18
114	13-methylbenzo[<i>b</i>]phenanthro [3 2- <i>d</i>]thiophene	0.0000	0.6581	2.2064	13.5075	22.6710	14.6758	511.77	507.74	508.66

1) Experimentally determined RI values (averaged) taken from Ref. [3] ; 2) The RI values predicted from QSRR Model 1. 3) The RI values predicted from QSRR Model 2.

2 结果与讨论

本文所研究的 114 种 PASHs 的气相色谱保留指数取自参考文献 [3]。

多元线性回归是一种经典的建模方法,本研究采用此法进行 PASHs 的 MEDV 参数与气相色谱保留值之间的相关性研究,删除 MEDV 10 个元素中全为 0 的 4 个元素,得到 6 个变量的模型(Model 1),结果见表 2。由 Model 1 所得预测值见表 1,预测值与实验值相关图见图 1。为检验 Model 1 的稳定性和预测能力,对模型进行留一法交互检验。所建立的 Model 1 的相关系数(R)为 0.994 7,交互检验相关系数(R_{cv})为 0.994 0,说明 Model 1 具有较好的稳定性和预测能力。

在多元线性回归分析中,可能存在变量间的偶然相关以及不利于模型稳定性的变量。为了消除这些影响,得到始终稳定的模型,需要在多元回归中引入逐步回归分析,按变量的重要性大小顺序引入方程,观察 R 、相对偏差(SD)、检验统计量(F)、 R_{cv} 、交互检验的相对偏差(SD_{cv})、交互检验的检验统计量(F_{cv})等随变量数 m 增加的变化,具体结果见表 2。

从表 2 可以看出,随着变量数 m 的增加, R 不断增加, SD 逐渐减小,这表明模型对样本内部的预测能力在不断增加。另外在留一法交互检验过程中,随着变量数 m 的增加, R_{cv} 不断增加,且 SD_{cv} 在变量数 m 为 6 时,和最小值(8.3606)很接近。结果表明,在 Model 1 构建过程中没有发生过度拟合现象,不需要进行变量的剔除。

表 2 逐步回归变量分析($N = 114$)
Table 2 Analysis of variables by SMR ($N = 114$)

m	a_0	a_8	a_6	a_5	a_3	a_2	a_1	R	SD	F	R_{cv}	SD_{cv}	F_{cv}
1	98.324		15.939					0.9562	21.8905	1196.5658	0.9547	22.2549	1154.0654
2	113.188	8.081	11.341					0.9814	14.4164	1453.0752	0.9803	14.8543	1365.4445
3	105.821	6.819	9.950	4.092				0.9860	12.5707	1286.0625	0.9850	13.0151	1197.2777
4	50.936	4.164	10.011	9.794	16.858			0.9942	8.1403	2338.4792	0.9937	8.4739	2155.9004
5	44.472	4.390	9.683	10.600	12.818	11.008		0.9946	7.9138	1980.8737	0.9940	8.3606	1772.5784
6	45.256	4.411	9.743	10.404	13.250	10.759	-149.812	0.9947	7.9062	1654.1061	0.9940	8.3814	1469.8853

$a_1, a_2, a_3, a_5, a_6, a_8$: the coefficients of the variables $V_1, V_2, V_3, V_5, V_6, V_8$, respectively ; CV : the results in cross validation ; m : variable amount.

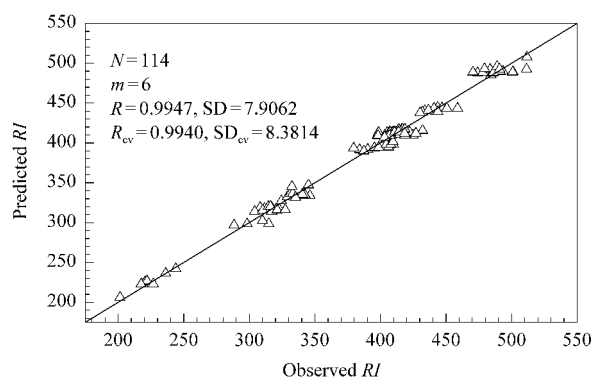


图 1 Model 1 中实验值与预测值的对比图

Fig. 1 Plot of observed versus predicted RI by Model 1

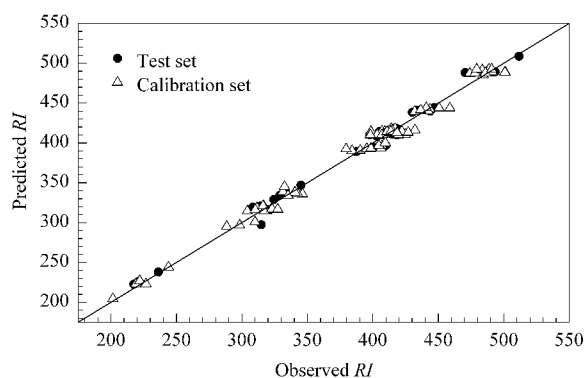


图 2 Model 2 中实验值与预测值的对比图

Fig. 2 Plot of observed versus predicted RI by Model 2

Tropsha 等^[16]最近的研究表明通过外部样本集能更准确地评价模型的预测能力。鉴于此,本研究中将样本集按 2:1 的比例分为校准集和测试集。首先,将表 1 中 3N ($N=1\sim 38$, 表 3 第 1 行) 序号的化合物选作测试集,然后利用校准集中的 76 种 PASHs 通过多元线性回归建立 QSRR 模型 (Model 2),并通过留一法交互检验进行模型验证,得到模型的回归方程为:

$$RI = 45.074 - 150.524V_1 + 12.925V_2 + 12.779V_3 + 10.165V_5 + 9.742V_6 + 4.683V_8$$

$$N = 76, R = 0.9946, SD = 8.0822, F = 1051.0144, R_{cv} = 0.9935, SD_{cv} = 8.8069, F_{cv} = 883.3604$$

通过 Model 2 对校准集和测试集进行预测,具体结果见表 1 和图 2。图 2 中预测值与实验值很好地吻合;且数据分布呈现一定的聚类趋势,这源于所收集的数据分布不均匀。另外,通过一元线性回归对测试集的实验值与预测值进行拟合,得到良好的相关系数和标准偏差(见表 3)。

同样的,本研究中还将序号 6N-1 或 6N (表 3 中第 2 行) 化合物,以及 6 组随机抽取 38 个化合物 (表 3 中第 3~8 行) 作为测试集,通过相同的数据处理方法,统计数据结果显示每个模型都有着优良的统计意义,校准集中实验值与预测值的相关系数都大于 0.994 (见表 3)。研究结果进一步表明所建立的 QSRR 模型具有优良的稳定性和预测能力。

PASHs 的 QSRR 研究工作一直受到人们的关注。Mössner 研究组^[8]采用如分子长度、厚度以及两者比例等比较简单的分子形状参数,研究参数与色谱保留值之间的线性关系,所得模型的相关系数不高,仅有 0.68 至 0.98。Can 等^[9]选用通过专业软件获得的结构参数、立体参数、电性参数、拓扑参数以及分子环境参数等作为模型构建参数,通过人工神经网络法建立了 6 个 PASHs 色谱保留值的预测模型,所得优化模型的预测能力良好,其交互检验相关系数达到 0.990;不过该模型仅选用了本文研究采用的 114 个数据中的 80 个数据,从而在一定程度上降低了模型的可信度。Schade 和 Anders-

表 3 不同测试集的统计分析结果

Table 3 Statistical results for different selection schemes of samples in the test set

No.	Calibration set			Test set		Sample number in the test
	R	SD	R_{cv}	R	SD	
1	0.9946	8.0822	0.9935	0.9953	6.2426	3, 6, 9, ..., 114 (3N; $N=1-38$)
2	0.9948	8.0709	0.9937	0.9941	7.1859	5, 6, 11, 121, 17, 181, ..., 114 (6N-1 or 6N; $N=1-19$)
3	0.9960	7.4247	0.9949	0.9888	9.7044	3, 13, 18, 23, 29, 32, 33, 37, 43, 44, 48, 50, 54, 56-60, 62, 64, 70, 76, 81, 83-89, 91, 92, 97-100, 104, 111, 114
4	0.9942	8.1227	0.9931	0.9951	6.4285	1-3, 9, 10, 17, 18, 20, 22, 24, 25, 28, 42, 43, 45, 52, 54-56, 58-60, 63, 65-67, 80, 82, 85, 86, 91, 95, 96, 99, 100, 102, 107, 114
5	0.9953	8.0557	0.9945	0.9915	8.5070	7, 19, 23, 24, 27, 29, 34, 38, 40, 45, 46, 51, 54-56, 58, 65, 66, 71, 73, 74, 76-80, 84, 86-89, 92, 94, 98, 100, 101, 103, 113, 114
6	0.9949	7.1797	0.9911	0.9936	7.5433	4, 5, 8, 9, 11, 12, 15, 17, 20, 23, 29, 36, 38, 39, 43, 55, 59, 70, 72, 75, 78, 79, 82, 83, 87, 91, 93, 95-97, 101-107, 113
7	0.9949	7.8554	0.9924	0.9939	7.1185	1, 5, 8, 13, 17, 22, 25, 29, 30, 33, 34, 36, 40-42, 44, 47, 62, 64-67, 70, 76, 79, 80, 83, 85, 87, 94, 95, 97, 98, 102, 104, 105, 111, 114
8	0.9953	7.7908	0.9943	0.9943	7.0033	1, 10, 13, 14, 16, 19-24, 28, 29, 31, 39, 44, 45, 48, 49, 51, 52, 55, 61, 65, 69, 74, 75, 78, 79, 81, 82, 85, 96, 100, 106, 107, 110, 111

son^[10]采用如分子中甲基取代位置等简单参数对 30 种烷基二苯并噻吩进行了 QSRR 研究,虽然所得模型的相关系数良好,但不适于预测结构有一定改变的 PASHs 的色谱保留指数。最近的相关研究工作是由 Zou 等^[11]开展的。他们采用量化计算的相关参数构建的模型具有很好的稳定性和预测能力,其交互检验相关系数到达 0.992。然而该模型构建过程需要大量的计算工作。与上述模型相比,本研究所建立的模型不仅有更好的稳定性和预测能力,而且参数获得容易;虽然该模型在扩展到含有 MEDV 方法中所定义的第 4 类原子(即一个非氢原子上接有 4 个非氢原子)的分子时显现出局限性,但鉴于目前发现绝大多数的 PASHs 不含有第 4 类原子,该模型依然适合在 PASHs 的 QSRR 研究中广泛应用。

3 结论

本文通过 MEDV 描述子表征了 114 种 PASHs 的结构信息,建立 MEDV 描述子与 PASHs 气相色谱保留值之间的 QSRR 模型,并通过内部以及外部双重验证的方法对模型的稳定性进行分析和验证。研究结果表明, MEDV 能较好地表征 PASHs 的结构信息,所建立的 QSRR 模型具有良好的稳定性和预测能力。MEDV 的计算易程序化,仅需要输入分子中的原子类型以及连接方式,其在实际中的应用更为简便。本文所建立的 QSRR 模型能为预测多环芳香硫化化合物的气相色谱保留指数提供一个便捷有效的方法,并为 PASHs 的分离、纯化、检测等方法

建立提供一定的理论支持。

参考文献:

- [1] Song C S. *Catal Today*, 2003, 86 : 211
- [2] Eastmond D A, Booth G M, Lee M L. *Arch Environ Contam Toxicol*, 1984, 13 : 105
- [3] Vassilaros D L, Kong R C, Later D W, et al. *J Chromatogr*, 1982, 252 : 1
- [4] Becker G, Colmsjo A. *Anal Chim Acta*, 1998, 376 : 365
- [5] Liang F Y, Lu M M, Birch M E, et al. *J Chromatogr A*, 2006, 1114 : 145
- [6] Zeng X L, Liu J, Liu J H, et al. *Chinese Journal of Analytical Chemistry* (曾小岚, 刘君, 刘建华, 等. 分析化学), 2006, 34(11) : 1546
- [7] Kaliszan R. *Chem Rev*, 2007, 107 : 3212
- [8] Mössner S G, de Alda M J L, Sander L C, et al. *J Chromatogr A*, 1999, 841 : 207
- [9] Can H, Dimoglo A, Kovalishyn V. *J Mol Struct : Theochem*, 2005, 723 : 183
- [10] Schade T, Andersson J T. *J Chromatogr A*, 2006, 1117 : 206
- [11] Xu H Y, Zou J W, Jiang Y J, et al. *J Chromatogr A*, 2008, 1198/1199 : 202
- [12] Liu S S, Liu Y, Li Z L, et al. *Acta Chimica Sinica* (刘树深, 刘堰, 李志良, 等. 化学学报), 2000, 58(11) : 1353
- [13] Liu S S, Liu Y, Yin D Q, et al. *Chinese Chem Lett*, 2005, 16(11) : 1559
- [14] Zhou Y, Sun L L, Mei H, et al. *Chromatographia*, 2006, 64 : 565
- [15] Li M P, Zhang S W, Zhao B, et al. *Journal of Analytical Science* (李美萍, 张生万, 赵冰, 等. 分析科学学报), 2008, 24(5) : 540
- [16] Tropsha A, Gramatica P, Gombar V K. *QSAR Comb Sci*, 2003, 22 : 69