2006 年第 64 卷 第 10 期, 1043～1050
化 学 学 报 ACTA CHIMICA SINICA
Vol. 64, 2006 No. 10, 1043～1050

• 研究论文 •

电拓扑状态预测有机磷酸酯类化合物的气相色谱保留指数

王 宇[a] 刘树深[a,b,*] 赵劲松[a] 王晓栋[a] 王连生[a]
([a]南京大学环境学院 污染控制与资源化研究国家重点实验室 南京 210093)
([b]同济大学环境学院 长江水环境教育部重点实验室 上海 200092)

摘要 以原子类型电拓扑状态指数(ETSI)有效表征 35 个有机磷酸酯类化合物(OP)的分子结构, 应用基于预测的变量选择与模型化(VSMP)方法建立 OP 化合物在 3 种不同固定相上的气相色谱保留指数(RI)与分子结构(ETSI)的定量相关模型. 结果表明, 影响不同固定相上 OP 色谱保留的主要结构因素都是由 7 个 ETSI 描述子对应的子结构碎片, 即: $=CH_2$, ≡C—, aaC—, =O, —O—, Cl 和 Br. 其中子结构 aaC—, =O 和—O 与 OP 化合物母体骨架密切相关, 而$=CH_2$, ≡C—, —Cl 和—Br 反映支链或取代基的变化. 通过多元线性回归法建立 OP 化合物在三个固定相上的定量结构-保留相关模型(QSRR)发现, 各 QSAR 模型的估计相关系数均在 0.99 以上, LOO 检验相关系数在 0.98 以上, 表明模型具有良好估计能力与稳定性. 应用 28 个 OP 训练集样本构建的 QSRR 模型预测外部 7 个检验集 RI 结果表明训练集模型具有良好预测能力.

关键词 电拓扑指数; 有机磷酸酯; 定量结构-保留相关; 基于预测的变量选择与模型化方法(VSMP)

Prediction of Gas Chromatographic Retention Indices of Organophosphates by Electrotopological State Index

WANG, Yu[a] LIU, Shu-Shen*[,a,b] ZHAO, Jin-Song[a]
WANG, Xiao-Dong[a] WANG, Lian-Sheng[a]
([a] *State Key Laboratory of Pollution Control and Resources Reuse*, *School of Environment*, *Nanjing University*, *Nanjing 210093*)
([b] *Key Laboratory of Yangtze Aquatic Environment*, *Ministry of Education*, *College of Environmental Science and Engineering*, *Tongji University*, *Shanghai 200092*)

Abstract Electrotopological state index (ETSI) for atom types was used to describe the structures of 35 organophosphates and a quantitative linear relationship between the ETSI descriptors and gas chromatographic retention indices (RI) was developed using the variable selection and modeling based on prediction (VSMP). It was found that some main structural factors influencing the RI of organophosphates are 7 substructures such as $=CH_2$, ≡C—, aaC— (where "a" refers to a chemical bond in the aromatic ring), =O, O, Cl and Br, which were related to the molecular skeleton of organophosphates, substituent groups on phenyl ring, and alkyls binding to the bond of P—O. Three best 7-variable models, with the calibrated correlation coefficient of $r>0.99$ and the validated correlation coefficient of $q>0.98$ for three stationary phases, were built by multiple linear regression, which shows a good estimation ability and stability of models. A prediction power for the external samples was validated by the model built from the training set with 28 organophosphates.

Keywords electrotopological state index; organophosphate; quantitative structure-retention relationship; variable selection and modeling based on prediction

* E-mail: ssliuhl@263.net or ssliu@nju.edu.cn
Received July 8, 2005; revised October 24, 2005; accepted January 25, 2006.
国家 973 计划(No. 2003CB415002)、国家自然科学基金(No. 20477018)、全国优秀博士学位论文作者基金(No. 200355)资助项目

有机磷酸酯类化合物(organophosphates, OP)广泛应用于医药制剂、农业化学品、可塑剂、阻燃剂和燃料添加剂以及军事工业上. 随着 20 世纪 70 年代对持久性有机氯杀虫剂的逐渐禁用, 高效低残留的有机磷酸酯类化合物逐渐成为杀虫剂的首选, 为迄今应用最为广泛的一类农药. 然而, 由于大量生产和广泛使用, 已对环境造成一定污染, 尤其对人类和非靶生物的毒害, 已引起科学工作者的高度重视[1~4].

为了解各种 OP 化合物的环境行为、毒性作用机制以及对生态系统的影响, 需要探索其结构与其物理化学性质或生物活性之间的相关关系. 定量分子结构-色谱保留相关(Quantitative structure-retention relationships, QSRR)是 QSAR 研究中的重要组成部分, 对于建立分子结构与色谱保留的变化规律, 估计与预测物质的物理化学性质与生物活性, 选择分离条件及深入探索色谱保留机理具有重要意义. 气相色谱保留指数是进行色谱定性分析的基础. 当固定相一定时, 化合物在色谱柱上的保留行为与该化合物的拓扑、几何和电性等结构特征密切相关. 借此, 可构建 QSRR 模型来实现不同结构特征化合物保留值的预测[5]. 然而, 目前这类化合物的 QSAR 研究很少[6,7]. 我们实验室曾应用 CoMFA 和 CoMSIA 方法对 OP 化合物对家蝇急性毒性的 QSAR 进行了较为深入的研究, 探讨了 OP 对乙酰胆酰酯酶的作用机理[8]. 原子类型电拓扑状态指数(ETSI)是一类基于分子二维拓扑结构与原子电性特征提出的理论分子描述子, 已在有机物理化性质与生物活性的评估与预测中得到较为广泛的应用[9~12], 国内应用很少[13].

为此, 本文选择 35 个 OP 化合物为研究对象, 以 ETSI 有效表征其分子结构, 应用基于预测的最佳子集回归(VSMP)方法[14,15]建立 OP 化合物结构与气相色谱保留指数之间的定量相关模型, 探讨 ETSI 方法预测 OP 化合物理化性质的可能性.

1 材料与方法

1.1 数据来源

选择文献[16]中 35 种 OP 化合物为研究对象(结构通式如图 1 所示). 它们在三种不同极性固定相(OV-101, DB-1701, DB-WX)上的气相色谱保留指数(RI)见表 1. 其中 OV-101 为非极性固定相, DB-1701 和 DB-WX 为极性固定相. OP 化合物的结构差异较大, 其中苯环上的间位和对位取代基(X)的电负性从给电子基团(—CH_3)变化到吸电子基团(—NO_2), 键合在 P—O 键上的烷基取代基(R)的体积从甲基变化到丁基. OP 的结构多样性以 ETSI 有效表征.

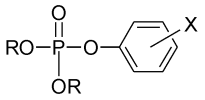

图 1 双烷基苯基磷酸酯化合物的分子骨架

Figure 1 Molecular skeleton for dialkylphenylphosphates

1.2 ETSI 计算

OP 化合物的分子结构采用原子类型电拓扑状态指数(Electrotopological State Indices for atom type, 简称 ETSI)进行描述. 电拓扑状态指数(E-状态指数)是分子连接性理论创始人 Kier 和 Hall 于 1990 年提出的基于原子水平的二维分子描述子, 已在药物设计与定量构效关系研究领域得到非常广泛的应用[9~11]. 原子类型 E-状态指数(ETSI)是 1995 年提出的基于原子类型的 E-状态指数[11,12]. E-状态指数编码各原子在分子中的拓扑环境及其与所有其它原子的电性相互作用, 其拓扑相关性基于原子之间的拓扑距离, 电性相互作用则基于原子固有状态以及其它原子对它的影响. ETSI 虽然与分子连接性密切相关, 但不同于分子连接性指数, 不必对分子图进行各个阶次的碎片子图分解, 而只需对各个非氢原子根据其在分子环境中的电子价态与局部拓扑信息进行分类.

根据我们先前研究的结果[14]并参考 ETSI 原始文献[12], 计算各化合物 ETSI 值的步骤总结如下:

首先计算各个原子类型的固有状态(I):

$$I=[(2/N)^2\square\delta^v+1]/\delta \quad (1)$$

$$(\delta=\sigma-h,\ \delta^v=\sigma+n+\pi-h)$$

式中 N 为该非氢原子价电子的主量子数, σ 和 π 分别为与该原子形成的 σ 和 π 键轨道电子数, h 为与之连接的氢原子数, n 为其孤对电子数, δ 反映其与相邻非氢原子形成的 σ 键数, δ^v 反映该原子的价电子状态. 显然, 某原子类型原子的固有状态描述该原子在分子中与其它原子的成键电子信息和邻接关系. 对于本文选择的 35 个 OP 化合物, 共有 12 种原子类型(常见有机分子中有 41 种原子类型[15]), 用符号表示为: sCH_3, $ssCH_2$, aaCH, stC, saaC, Tn, sddN, dO, ssO, ssssdP, sCl, sBr. 它们对应的原子类型编号(No.)、子结构、统一描述子(descriptor)以及 ETSI 符号如表 2 所示. 表中大写 S 代表 ETSI, 小写 s, d 和 t 分别表示单键、双键和三键, 小写 a 表示芳环中的键.

然后计算分子拓扑中其它非氢原子的扰动或影响(ΔI), 得到 E-状态指数(S), 最后将属于相同原子类型的各非氢原子的 E-状态指数相加即得到分子中各原子类型的 ETSI:

表 1　35 个 OP 的结构及在不同固定相上的气相色谱保留指数(RI)

Table 1　Structures of 35 OPs and their retention indices (RI) on three stationary phases

No.	R	X	RI (OV-101)		RI (DB-1701)		RI (DB-WX)	
			Observed	Estimated	Observed	Estimated	Observed	Estimated
1	CH_3	H	1420	1376	1680	1637	2140	2127
2	CH_3	3-CH_3	1488	1508	1758	1753	2219	2226
3	CH_3	4-CH_3	1504	1514	1768	1759	2245	2230
4	CH_3	4-OCH_3	1637	1650	1840	1812	2350	2399
5	CH_3	3-Cl	1560	1555	1816	1809	2300	2312
6	CH_3	4-Cl	1574	1568	1830	1821	2314	2320
7	CH_3	3-Br	1652	1659	1910	1913	2392	2395
8	CH_3	4-Br	1666	1669	1924	1922	2410	2401
9	CH_3	3-CN	1678	1686	2003	2031	2625	2618
10	CH_3	4-CN	1706	1705	2041	2049	2662	2634
11	CH_3	3-NO_2	1795	1807	2065	2156	2720	2734
12	CH_3	4-NO_2	1810	1827	2085	2172	2737	2747
13	CH_2CH_3	H	1502	1483	1756	1725	2210	2223
14	CH_2CH_3	4-CH_3	1617	1621	1847	1848	2310	2326
15	CH_2CH_3	3-Cl	1667	1657	1891	1894	2390	2406
16	CH_2CH_3	4-Cl	1675	1671	1912	1906	2410	2415
17	CH_2CH_3	3-Br	1777	1766	1996	2002	2490	2492
18	CH_2CH_3	4-Br	1790	1776	2010	2011	2510	2498
19	CH_2CH_3	3-CN	1780	1785	2096	2113	2725	2713
20	CH_2CH_3	4-CN	1800	1805	2140	2132	2762	2729
21	CH_2CH_3	3-NO_2	1880	1893	2264	2226	2820	2819
22	CH_2CH_3	4-NO_2	1895	1917	2284	2246	2838	2834
23	CH_2CH_3	2,4-Cl	1837	1832	2031	2033	2490	2487
24	CH_2CH_3	2,5-Cl	1844	1823	2044	2024	2502	2482
25	$CH_2CH_2CH_2CH_3$	H	1889	1910	2103	2122	2596	2548
26	$CH_2CH_2CH_2CH_3$	3-CH_3	2025	2044	2202	2240	2614	2649
27	$CH_2CH_2H_2CH_3$	4-CH_3	2047	2050	2224	2245	2630	2652
28	$CH_2CH_2CH_2CH_3$	4-OCH_3	2183	2178	2239	2289	2861	2818
29	$CH_2CH_2CH_2CH_3$	3-Cl	2065	2079	2270	2285	2730	2728
30	$CH_2CH_2CH_2CH_3$	4-Cl	2085	2094	2295	2298	2774	2738
31	$CH_2CH_2CH_2CH_3$	4-Br	2190	2204	2417	2408	2810	2825
32	$CH_2CH_2CH_2CH_3$	3-CN	2215	2202	2515	2501	2989	3036
33	$CH_2CH_2CH_2CH_3$	4-CN	2228	2224	2552	2521	3020	3052
34	$CH_2CH_2CH_2CH_3$	3-NO_2	2326	2290	2637	2595	3112	3128
35	$CH_2CH_2CH_2CH_3$	4-NO_2	2345	2320	2674	2619	3180	3145

$$\Delta I_i=\sum_{j\neq i}^{\text{all}}(I_i-I_j)/(d_{ij}+1)^2 \quad (2)$$

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

式中 d_{ij} 为第 i 和第 j 原子之间的拓扑距离. ΔI 即反映了分子拓扑环境中其它不同电特征的非氢原子对固有状态 I 的影响, 这也正是命名为电拓扑状态指数的缘由.

表 2 12 种原子类型及相应编号、子结构、描述子与 ETSI 符号

Table 2 12 atomic types and their coding number, substructure, descriptor, and ETSI symbol

No.	Atom type	Substructure	Descriptor	ETSI
1	sCH_3	—CH_3	x_1	$SsCH_3$
3	$ssCH_2$	=CH_2	x_3	$SsssCH_2$
6	aaCH	aCHa	x_6	SaaCH
9	stC	≡C—	x_9	SstC
11	saaC	aaC—	x_{11}	SsaaC
18	tN	≡N	x_{18}	StN
22	sddN	⩾N=	x_{22}	SsddN
25	dO	=O	x_{25}	SdO
26	ssO	—O—	x_{26}	SssO
32	sssdP	≡P=	x_{32}	SssssdP
39	sCl	—Cl	x_{39}	SsCl
40	sBr	—Br	x_{40}	SsBr

1.3 描述变量优化选择

35 个 OP 化合物共有 12 个非零 ETSI 描述子. 然而, 影响 OP 化合物保留指数变化的描述子可能只是其中的一部分, 另一方面, 最佳模型中包含 12 个描述子也太多. 因此, 必须进行变量优化筛选. 目前流行的方法有逐步回归分析、遗传算法以及神经网络等. 然而, 其中许多方法是以模型估计统计量为目标的, 常常只能说明模型对内部样本的估计能力. 因此, 本文采用我们先前开发的基于预测的变量选择与模型化(VSMP)方法[14,15]进行变量优化选择. 首先从各种变量组合(称子集)中系统地选择一个子集, 进行多元回归建模, 如果模型相关系数大于初始给定的变量自相关系数阀值, 则进行交互检验(否则选择下一个子集), 若其检验相关系数大于给定值, 则替代给定值(否则选择下一个子集); 然后选择下一个子集进行同样的过程, 直到全部子集回归与检验完成为止[17].

2 结果与讨论

2.1 ETSI 对 OP 化合物结构的分辨

研究表明, 在 12 个原子类型 E-状态指数 ETSI 张成的模式空间中[17], 35 个 OP 化合物是相互分离的, 两两之间的欧氏距离都大于零, 这表明结构不同的化合物具有不同的电拓扑状态矢量, 也即 ETSI 可以完全分辨 35 个 OP 化合物, 其结构分辨率为 100%. 同时研究表明, OP 化合物的多样性或相似性可在 ETSI 构成的模式空间中充分描述. 例如, 距离最小的 26 和 27 号化合物(d=0.0099), 其结构最相似, 只有—CH_3 取代位的微小差异. 又如结构差异很大的 10 和 35 号化合物之间的模式距离则达到 d=591.9895.

2.2 VSMP 选择最优变量

以 12 个非零 ETSI 描述子为自变量(x), 分别以不同极性固定相上 OP 化合物的保留指数(RI)为因变量(Y)构建数据集, 应用基于预测的最佳子集回归变量选择与模型化方法(VSMP)建立不同固定相上保留指数的最佳子集回归模型. 研究结果表明, 当变量数 m 从 1, 2, 3 变化到 8 时, OP 在三个固定相(OV-101, DB-1701 和 DB-WX)上保留指数 QSRR 模型的均方根误差(RMS)随变量数的变化情况如图 2 所示. 由图可知, 三个不同固定相上的最佳变量子集都是 7 个 ETSI 描述子 x_3, x_9, x_{11}, x_{25}, x_{26}, x_{39} 和 x_{40}(具体数值参见表 3)的组合. 这些变量之间的自相关系数除 $r(x_{11}, x_{25})$=−0.6098 与 $r(x_3, x_{26})$=0.5472 外其它各对 ETSI 之间的相关系数都小于 0.21, 表明最佳模型中不存在假相关.

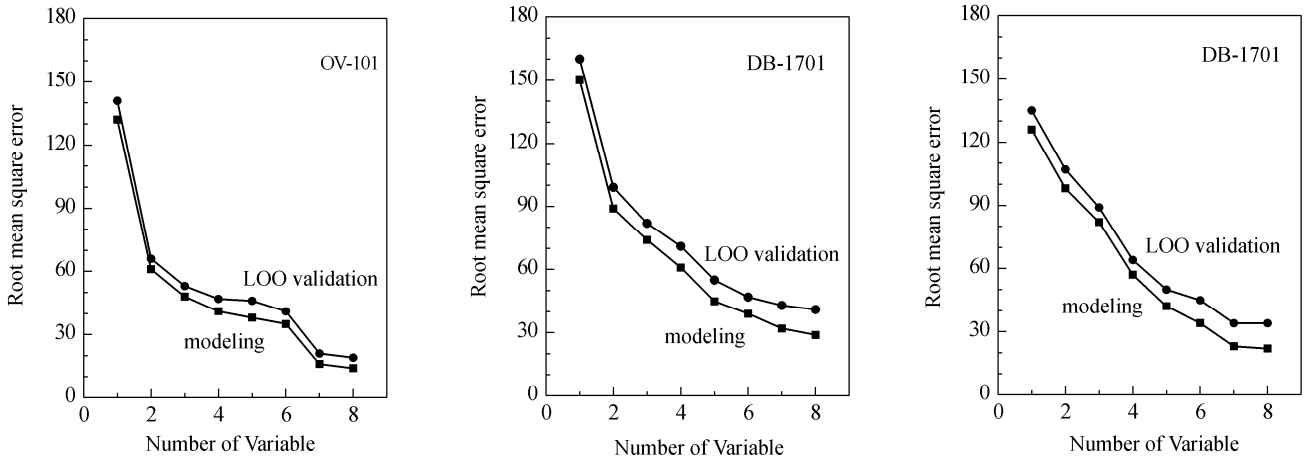

图 2 三个固定相上 OP 化合物模型均方根误差随变量数的变化曲线

Figure 2 Plot of the root mean square error versus the number of variables

表 3　35 个 OP 化合物的 7 个 ETSI 描述子

Table 3　The values of 7 ETSI descriptors for 35 OPs

No.	x_3	x_9	x_{11}	x_{25}	x_{26}	x_{39}	x_{40}
1	0.0000	0.0000	0.4533	11.4534	14.2153	0.0000	0.0000
2	0.0000	0.0000	1.4847	11.5647	14.3759	0.0000	0.0000
3	0.0000	0.0000	1.5600	11.5383	14.3308	0.0000	0.0000
4	0.0000	0.0000	1.0787	11.5766	19.2748	0.0000	0.0000
5	0.0000	0.0000	0.8250	11.5216	14.2053	5.7087	0.0000
6	0.0000	0.0000	0.9478	11.5054	14.2062	5.6635	0.0000
7	0.0000	0.0000	1.2503	11.5494	14.3153	0.0000	3.2622
8	0.0000	0.0000	1.3425	11.5266	14.2866	0.0000	3.2721
9	0.0000	1.9378	0.6736	11.5701	14.1879	0.0000	0.0000
10	0.0000	1.9579	0.8132	11.5429	14.1886	0.0000	0.0000
11	0.0000	0.0000	−0.1272	32.5173	13.9627	0.0000	0.0000
12	0.0000	0.0000	0.0785	32.2791	14.0151	0.0000	0.0000
13	0.5637	0.0000	0.4748	11.9159	15.1637	0.0000	0.0000
14	0.5618	0.0000	1.5889	12.0008	15.2792	0.0000	0.0000
15	0.4957	0.0000	0.8564	11.9841	15.1536	5.7765	0.0000
16	0.5096	0.0000	0.9767	11.9679	15.1545	5.7181	0.0000
17	0.5382	0.0000	1.2818	12.0119	15.2636	0.0000	3.2963
18	0.5433	0.0000	1.3714	11.9891	15.2349	0.0000	3.2994
19	0.4349	1.9658	0.7051	12.0326	15.1363	0.0000	0.0000
20	0.4594	1.9802	0.8420	12.0054	15.1369	0.0000	0.0000
21	0.2901	0.0000	−0.0957	33.2045	14.9111	0.0000	0.0000
22	0.3418	0.0000	0.1074	32.9269	14.9635	0.0000	0.0000
23	0.6108	0.0000	3.1622	12.1868	15.6577	0.0000	0.0000
24	0.6158	0.0000	3.0545	12.2169	15.7197	0.0000	0.0000
25	4.3736	0.0000	0.4990	12.4656	16.0695	0.0000	0.0000
26	4.3679	0.0000	1.5534	12.5769	16.2300	0.0000	0.0000
27	4.3683	0.0000	1.6232	12.5505	16.1849	0.0000	0.0000
28	4.2513	0.0000	1.1419	12.5888	21.2351	0.0000	0.0000
29	4.2076	0.0000	0.8937	12.5338	16.0594	5.8759	0.0000
30	4.2391	0.0000	1.0110	12.5175	16.0603	5.8006	0.0000
31	4.3224	0.0000	1.4057	12.5388	16.1407	0.0000	3.3405
32	4.0544	2.0064	0.7423	12.5823	16.0421	0.0000	0.0000
33	4.1108	2.0137	0.8764	12.5550	16.0427	0.0000	0.0000
34	3.6945	0.0000	−0.0585	34.0948	15.8168	0.0000	0.0000
35	3.8137	0.0000	0.1417	33.7642	15.8692	0.0000	0.0000

结合表 2 中 ETSI 描述子与子结构的关系可知, 它们分别对应 OP 化合物分子中的=CH_2, ≡C—, aaC—, =O, —O—, —Cl 和—Br 等结构碎片. 这表明 OP 化合物分子骨架中的三个子结构(aaC—, =O 和 O)以及支链或取代基中的四个子结构(=CH_2, ≡C—, —Cl 和—Br)是影响色谱保留指数的最主要因素. 为便于理解, 图 3 给出了第 6 号 OP 化合物中 4 种原子类型(11, 25, 26 和 39)与子结构的关系.

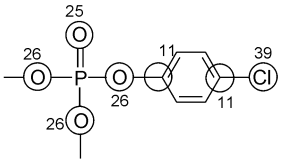

图 3　第 6 号 OP 化合物中几种原子类型与子结构之间的关系

Figure 3　Relationships between some atomic types and sub-structures of compound of No. 6

通过 VSMP 选择的 7 个优化 ETSI 描述子是否还能分辨 35 个 OP 化合物? 结果表明, 在 7 个 ETSI 张成的

模式空间中, 35 个 OP 化合物仍然是相互分离的, 任意两个 OP 之间的欧氏距离都大于零, 结构上只是 X 取代基(—CH_3)在苯环间位与对位差异的第 26 和 27 号 OP 之间的距离仍然最短($d_{26,27}$=0.0076), 这充分说明 7 个优化 ETSI 也能完全分辨 35 个 OP 化合物的结构多样性. 同时, 这也说明了 VSMP 选择优化变量的可靠性.

2.3 模型与检验

应用多元线性回归(MLR)方法, 建立 7 个最佳 ETSI 描述子与 OP 化合物在不同固定相上的保留指数定量相关模型(QSRR):

$$\mathrm{RI}=b_0+b_3\square x_3+b_9\square x_9+b_{11}\square x_{11}+b_{25}\square x_{25}+b_{26}\square x_{26}+b_{39}\square x_{39}+b_{40}\square x_{40} \quad (4)$$

三个 QSRR 模型的回归系数(b_i)与相应标准偏差(±后的数据)、样本数(n)、估计相关系数(r)与均方根误差(RMSE)、LOO 检验相关系数(q)与均方根误差(RMSV)、Fisher 统计量(F)结果见表 4. 由表可知, 各个 QSRR 模型都具有良好的估计与 LOO 交互检验统计性. 在固定相 OV-101, DB-1701 和 DB-WX 上的模型估计相关系数(r)分别为 0.9980, 0.9921 和 0.9961, 相应 LOO 检验相关系数(q)分别为 0.9965, 0.9859 和 0.9916. 表明各模型具有良好的估计能力与稳定性. 由三个不同固定相 QSRR 模型估计的 OP 化合物保留指数(见表 1 中“Estimated”栏)与原实验观测保留指数(见表 1 中“Observed”栏)的整体相关情况由图 4 给出. 由表 1 可知, 对于 DB-1701 固定相, 由 QSRR 模型估计的第 11 和第 12 号 OP 化合物的误差偏大, 分别达到 91 和 87, 大大超出了标准偏差的 2 倍, 从而导致总体估计均方根误差偏大(RMSE=32). 如果去除这 2 个化合物后重新计算, 得 RMSE=24, 均方根误差大大降低.

一个好的 QSRR 模型除了具有良好校正统计性与稳定性之外, 还必须具有对外部样本的良好预测能力. 为此, 将 35 个化合物分为训练集和检验集, 以训练集建立模型, 然后预测检验集样本. 按保留指数大小排序后均匀选择 28 个 OP 化合物作为训练集样本, 余下的 7 个 OP 作为外部检验集样本. 同样应用 VSMP 方法, 以 LOO 检验均方根误差为优化目标, 对训练集进行不同变量数的最佳变量选择. VSMP 分析结果表明, 除 OV-101 固定相外, 选择的最佳变量子集与应用全部 35 个化合物所得结果相同, 即仍是 7 个 ETSI 描述子 x_3, x_9, x_{11}, x_{25}, x_{26}, x_{39} 和 x_{40} 的组合, 这说明 28 个训练集与 35 个 OP 的整体数据集具有相同的 QSRR 规律. 三个训练集模型的有关统计量与全部 35 个 OP 化合物建立的模型统计量之间没有显著性差异, 说明训练集模型同样具有良好估计能力与稳定性. 而且也具有良好的对外部样本的预测能力.

应当指出, 对于 OV-101 固定相, 最佳变量组合虽然也是 7 变量组合, 但与另两个模型最佳子集不同的是描述子 x_9 变成了 x_{18}. 这是什么原因? 分析这两个变量之间的自相关系数发现, 其自相关系数达到 0.9999, 高度相关. 由此可以认为以 x_9 替代 x_{18}, 模型统计性不应有显著性变化. 应用 MLR 方法分析与比较含 x_9 和 x_{18} 时模型的统计量:

表 4 OP 化合物在不同固定相上的保留指数相关模型及有关统计量

Table 4 Models for predicting the retention indices of OPs on three stationary phases

Statistic	OV-101	DB-1701	DB-WX
b_0	(493.218±47.714)	(973.853±96.409)	(1121.72±70.33)
b_3	(98.8053±2.0289)	(94.4854±4.0996)	(70.2066±2.9905)
b_9	(145.630±5.1668)	(190.032±10.440)	(242.317±7.616)
b_{11}	(118.926±5.9729)	(106.219±12.069)	(86.0354±8.8038)
b_{25}	(24.1771±0.6069)	(27.8357±1.2264)	(31.7164±0.8946)
b_{26}	(38.8447±2.7783)	(20.8178±5.6139)	(42.4047±4.0951)
b_{39}	(23.3571±1.7236)	(23.0379±3.4827)	(26.4565±2.5405)
b_{40}	(55.8445±3.0071)	(57.3720±6.0761)	(58.8996±4.4323)
n	35	35	35
r^2	0.9960	0.9843	0.9922
r	0.9980	0.9921	0.9961
RMSE	16	32	23
F	959.74	242.50	493.17
q^2	0.9931	0.9720	0.9833
q	0.9965	0.9859	0.9916
RMSV	21	43	34

n＝28, m＝7, r＝0.9980, RMSE＝16, F＝702.69,
q＝0.9954, RMSV＝24　　(含 x_9 不含 x_{18})
n＝28, m＝7, r＝0.9980, RMSE＝16, F＝704.51,
q＝0.9955, RMSV＝24　　(含 x_{18} 不含 x_9)

很明显两者统计量之间几乎没有差异. 说明三个固定相模型的最佳子集是相同的. 应用 28 个训练集样本模型估计与预测的保留指数和实验保留指数之间的总体相关情况见图 5. 计算表明误差较大的化合物仍是 DB-1701 固定相模型估计的第 11 与 12 号 OP 以及 DB-WX 固定相模型预测的第 4 号检验化合物, 误差分别为 92, 87 和 82.

3　结论

原子类型 E-状态指数(ETSI)能有效表征 35 个有机磷酸酯类化合物的分子结构. 通过 VSMP 优化选择的 7 个 ETSI 描述子与 OP 化合物在不同固定相上的保留指数具有高度相关性. ETSI 与保留指数之间的定量相关模型不仅对内部样本具有良好的估计能力, 同时对外部样本具有良好的预测能力.

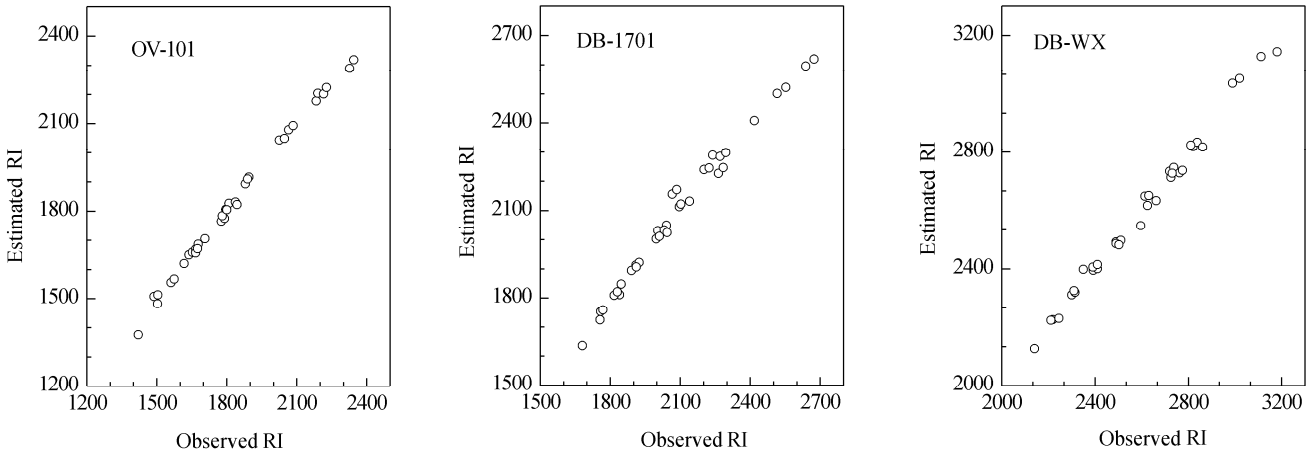

图 4　35 个 OP 在不同固定相上的 QSRR 模型估计与实验保留指数相关图
Figure 4　Plot of estimated by QSRR versus observed RI of 35 OPs on various stationary phases

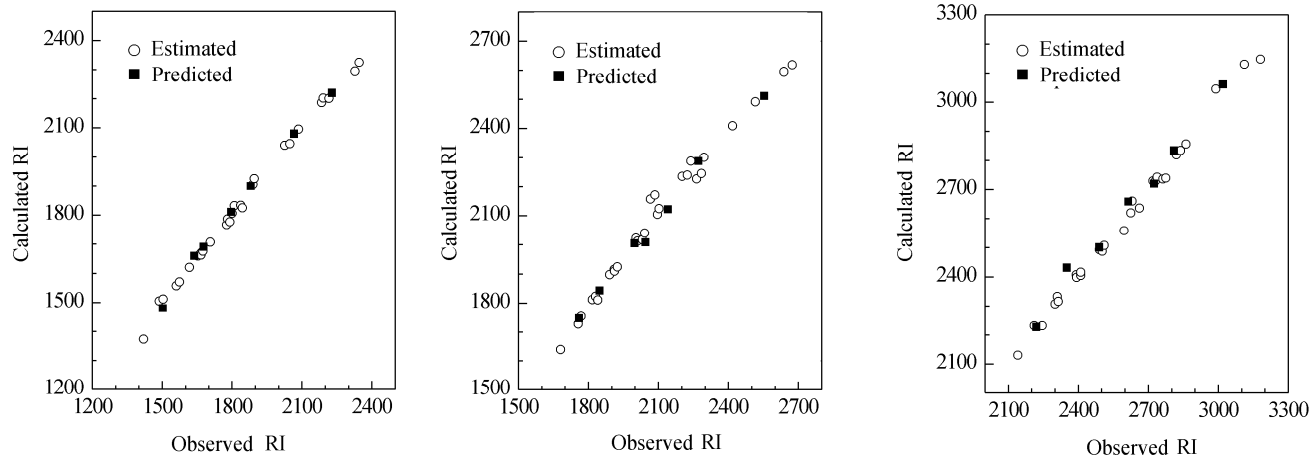

图 5　28 个训练集 OP 在不同固定相上的 QSRR 模型估计及预测与实验保留指数相关图
Figure 5　Plot of estimated versus observed RI of 28 OPs in training set on various stationary phases

References

1　Rickwood, C. J.; Galloway, T. S. *Aquat. Toxicol.* **2004**, *67*, 45.
2　Solberg, Y.; Belkin, M. *Curr. Awaren.* **1997**, *18*, 183.
3　Parran, D. K.; Magnin, G.; Li, W.; Jortner, B. S.; Ehrich, M. *Neurotoxicology* **2005**, *26*, 77.
4　Silva, D.; Cortez, C. M.; Cunha-Bastos, J.; Louro, S. R. W. *Toxicol. Lett.* **2004**, *147*, 53.
5　Liu, S. S.; Yin, C. S.; Wang, L. S. *Chin. Chem. Lett.* **2002**, *13*, 791.
6　Singh, A. K. *SAR QSAR Environ. Res.* **2001**, *12*, 275.

7 Drew, M. G. B.; Lumley, J. A.; Price, N. R. *Quant. Struct.-Act. Relat.* **1999**, *18*, 573.
8 Zhao, J. S.; Wang, B.; Dai, Z. X.; Wang, X. D.; Kong, L. R.; Wang, L. S. *Chin. Sci. Bull.* **2004**, *49*, 65 (in Chinese).
(赵劲松, 王斌, 戴朝霞, 王晓栋, 孔令仁, 王连生, 科学通报, **2004**, *49*, 65.)
9 Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
10 Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76.
11 Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*, Academic Press, New York, **1999**, pp. 67～75.
12 Hall, L. H.; Kier, L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039.
13 Liu, X. H.; Luo, W. R.; Wang, L. S. *Chin. Environ. Sci.* **2004**, *24*, 442 (in Chinese).
(刘新会, 骆文茹, 王连生, 中国环境科学, **2004**, *24*, 442.)
14 Liu, S. S.; Yin, D. Q.; Cui, S. H.; Wang, L. S. *Chin. J. Chem.* **2005**, *23*, 622.
15 Liu, S. S.; Liu, H. L.; Yin, C. S.; Wang, L. S. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 964.
16 Gandhe, B. R.; Prasad, P. R.; Danikhel, R. K.; Shinde, S. K.; Srivastava, R. K.; Batra, B. S.; Rao, K. M. *Pestic. Sci.* **1990**, *29*, 379.
17 Liu, S. S. *Structural Characterization of Organic Compounds by the Molecular Electronegativity Distance Vector* (*MEDV*), Higher Education Press, Beijing, **2005**, pp. 50～59.
(刘树深, 有机物分子电性距离矢量表征及其应用, 高等教育出版社, 北京, **2005**, pp. 50～59.)

(A0507084 QIN, X. Q.; LING, J.)