

硝基苯类化合物对四膜虫毒性的 QSAR 研究

杜浩¹, 杨静¹, 王猛², 崔世海¹

(1. 南京师范大学化学与环境科学学院, 江苏 南京 210097)

(2. 江苏省海安县环境保护局, 江苏 海安 226600)

[摘要] 以电拓扑状态指数表征硝基苯类化合物的分子结构, 计算了 42 个硝基苯类化合物中所含 41 种原子类型的电拓扑指数. 建立优化电拓扑指数与硝基苯化合物对四膜虫的半抑制生长浓度负对数值之间的 QSAR 模型. 结果表明所建模型具有良好的拟合效果和稳健度. 通过建立的 QSAR 模型对检验集数据的测试, 表明该模型有很好的预测能力.

[关键词] 硝基苯类化合物, 电拓扑状态指数, QSAR

[中图分类号] X503.2 [文献标识码] A [文章编号] 1001-4616(2009)04-0068-04

QSAR Study of Nitrobenzene Compounds Toxicity to the Tetrahymena Pyriformis

Du Hao¹, Yang Jing¹, Wang Meng², Cui Shihai¹

(1. School of Chemistry and Environmental Science, Nanjing Normal University, Nanjing 210097, China)

(2. Hainan Environmental Protection Bureau, Hainan 226600, China)

Abstract The electrotopological state index (ETSI) was employed to characterize the structures of 42 nitrobenzene compounds. 41 ETSI indices of the compounds were calculated. The method for variable selection and modeling based on prediction was used to select an optimal combination of the variables from 41 ETSI descriptors. QSAR model was constructed between the negative logarithmic value of the half inhibitory concentration of the nitrobenzene compounds to *Tetrahymena Pyriformis* and molecular structures of nitrobenzene compounds. The results indicated that the model presented good estimated ability and stability. The predictive power for external compounds was validated by QSAR model constructed by the training set.

Key words nitrobenzene compounds, electrotopological state index, QSAR

硝基苯类化合物是重要的化工原料, 在染料、香料、炸药等方面有着广泛的应用, 然而环境中该类污染物的存在, 对人类及野生动物的健康产生严重的危害. 据报道硝基苯类化合物能够使血液中的血红蛋白变成氧化血红蛋白(即高铁血红蛋白), 大大阻止了血红蛋白的输送氧的作用, 因而使人及动物呈现呼吸急促和皮肤苍白的现象, 长久中毒以后, 则会引起肝脏损坏等病症. 我国已经将硝基苯、对硝基甲苯、2,4-二硝基甲苯、三硝基甲苯、对硝基氯苯、2,4-二硝基氯苯等 6 种化合物列为水体中的优先控制污染物.

从结构分析, 硝基苯类化合物中苯环上的硝基为一强吸电子取代基团, 而这些潜在的亲电反应可能导致急性毒性、突变性及皮肤过敏作用, 主要是因为硝基芳香衍生物与细胞核酸或蛋白中的亲核试剂发生反应所引起. 取代基团的性质不同可以产生许多不同的亲电反应, 根据这一性质, 众多科研人员对硝基苯类化合物的急性毒性建立了相关定量结构-活性相关(Quantitative Structure-Activity Relationship, 简称 QSAR)模型. 徐镜波^[1]等基于 20 种硝基化合物的辛醇/水分配系数 $\lg P$ 、一阶价分子连接性指数等参数对鲤鱼鳃和肾的 ATPase 活性的半抑制浓度进行构效分析, 认为该类化合物的毒性与最低占有轨道能量、取代基常数及苯环上的硝基数目和位置直接相关. 王连生等^[2]应用三维结构描述子对 25 种硝基苯类化合物进行 QSAR 研究, 结果表明该类化合物的毒性与苯环取代基的种类、位置及数目相关, 产生毒性的机制为反应

收稿日期: 2009-01-10

基金项目: 江苏省高校自然科学基金(07KJB610061)资助项目.

通讯联系人: 崔世海, 副教授, 研究方向: 有机污染化学. E-mail: cuishihai@njnu.edu.cn

型,同时构建了具有一定预测能力的 QSAR 模型. 本文基于二维描述子电拓扑状态指数 (Electrotopological State Indices, 简称 ETSI) 表征 42 种硝基苯类化合物的结构, 对四膜虫的半抑制生长浓度进行结构活性相关研究, 分析影响该类化合物毒性的重要分子结构单元, 所构建的模型具有更优越的拟合和预测能力.

1 材料与方法

1.1 活性数据

硝基苯的毒性数据来自文献 [3], 48 h 四膜虫生长半抑制浓度负对数值 $\lg(IGC_{50}^{-1})$ 数据见表 1, 其中浓度单位为 mg/L.

表 1 硝基苯类化合物活性数据实验值、计算值以及部分结构描述子

Table 1 The observed values, calculated values and six ETSI descriptors of nitrobenzene compounds

编号	化合物	$\lg(IGC_{50}^{-1})$ 活性数据			化合物的部分结构描述子					
		实验	拟合	预测	x_1	x_6	x_{11}	x_{22}	x_{25}	x_{39}
1	2,6-二甲硝基苯	0.30	0.47	0.19	3.482 4	5.299 5	1.664 5	-0.342 6	20 896 3	0.000 0
2	2,3-二甲硝基苯*	0.56	0.56	0.54	3.620 8	5.074 5	1.904 5	-0.358 8	20 759 1	0.000 0
3	2-甲基-3-氯硝基苯	0.68	0.75	0.81	1.625 4	4.623 5	1.010 3	-0.443 3	20 641 8	5.653 4
4	2-甲基硝基苯	0.052	0.24	0.18	1.720 4	6.653 7	0.884 4	-0.379 6	20 454 6	0.000 0
5	2-氯硝基苯	0.68	0.43	0.60	1.588 4	6.070 8	0.122 1	0.000 0	19 399 1	4.264 2
6	2-甲基-5-氯硝基苯	0.82	0.75	0.59	1.675 0	4.598 4	1.082 8	-0.443 3	20 641 8	5.556 4
7	2,4,5-三氯硝基苯	1.53	1.55	1.53	0.000 0	2.356 9	0.056 3	-0.620 5	20 609 3	16.598 1
8	2,5-二氯硝基苯*	1.13	1.14	1.17	0.000 0	4.100 0	0.236 0	-0.575 2	20 472 9	10.988 6
9	6-氯-1,3-二硝基苯	1.98	1.85	1.66	0.000 0	3.051 0	-0.928 0	-1.487 5	41 044 0	5.431 6
10	硝基苯	0.14	0.01	0.14	0.000 0	7.933 8	0.136 6	-0.416 7	20 013 0	0.000 0
11	3-甲基硝基苯*	0.054	1.31	0.45	1.826 4	6.521 3	1.064 2	-0.395 8	20 317 4	0.000 0
12	1,3-二硝基苯	0.89	1.02	1.27	0.000 0	4.593 5	-0.547 6	-1.347 2	40 634 8	0.000 0
13	3,4-二氯硝基苯	1.16	0.26	1.13	0.000 0	3.902 9	0.456 3	-0.525 6	20 336 5	11.052 2
14	4-甲基硝基苯	0.17	0.15	0.14	1.892 6	6.431 4	1.177 2	-0.403 3	20 235 6	0.000 0
15	1,4-二硝基苯*	1.30	1.11	0.90	0.000 0	4.381 4	-0.304 3	-1.214 8	40 471 1	0.000 0
16	4-氯硝基苯	0.43	0.51	0.42	0.000 0	5.698 4	0.565 0	-0.462 0	20 149 4	5.493 7
17	2,3,5,6-四氯硝基苯*	1.82	1.78	1.66	0.000 0	1.250 7	-0.880 7	-0.733 7	20 932 8	22.208 7
18	6-甲基-1,3-二硝基苯*	0.87	1.14	1.50	1.529 6	3.549 5	-0.081 2	-1.296 9	41 299 1	0.000 0
19	3-氯硝基苯	0.73	0.52	0.59	0.000 0	5.835 7	0.404 5	-0.480 3	20 200 1	5.484 4
20	1,2-二硝基苯	1.25	1.39	1.18	0.000 0	4.950 9	-0.967 4	-1.559 3	40 909 2	0.000 0
21	2-溴硝基苯*	0.75	0.82	0.94	0.000 0	6.446 6	0.613 6	-0.426 5	20 394 6	0.000 0
22	6-溴-1,3-二硝基苯	2.31	2.08	2.03	0.000 0	3.372 4	-0.382 0	-1.364 6	41 208 4	0.000 0
23	3-溴硝基苯	1.03	0.82	0.38	0.000 0	6.277 7	0.829 8	-0.425 8	20 275 7	0.000 0
24	4-溴硝基苯*	0.38	0.61	0.76	0.000 0	6.171 0	0.959 7	-0.424 2	20 204 9	0.000 0
25	2,4,6-三甲硝基苯	0.86	0.73	0.56	5.458 3	3.649 0	2.769 9	-0.329 3	21 118 9	0.000 0
26	5-甲基-1,2-二硝基苯	1.52	1.61	1.48	1.635 6	3.774 5	-0.321 2	-1.525 1	41 436 3	0.000 0
27	2,4-二氯硝基苯	0.99	1.03	1.20	0.000 0	4.025 2	0.318 9	-0.556 9	20 422 2	11.012 9
28	3,5-二氯硝基苯	1.13	1.03	0.94	0.000 0	3.927 5	0.440 9	-0.543 9	20 387 3	11.010 6
29	6-碘-1,3-二硝基苯	2.12	1.85	1.72	0.000 0	3.051 0	-0.928 0	-1.487 5	41 044 0	5.431 6
30	2,3,4,5-四氯硝基苯	1.78	2.07	2.02	0.000 0	1.053 6	-0.660 4	-0.684 1	20 796 5	22.272 3
31	2,3-二氯硝基苯*	1.07	1.07	1.07	0.000 0	4.275 2	0.013 4	-0.575 2	20 472 9	11.036 0
32	2,5-二溴硝基苯	1.37	1.62	1.63	0.000 0	4.808 9	1.261 8	-0.435 7	20 657 4	0.000 0
33	1,2-二氯-4,5-二硝基苯	2.21	2.26	2.28	0.000 0	2.064 9	-0.406 5	-1.634 0	41 278 0	9.586 6
34	3-甲基-4-溴硝基苯	1.16	1.09	1.02	1.816 5	4.675 2	1.896 5	-0.403 3	20 509 4	0.000 0
35	2,3,4-三氯硝基苯*	1.51	1.45	1.40	0.000 0	2.532 2	-0.166 4	-0.620 5	20 609 3	16.645 5
36	2,4,6-三氯硝基苯	1.43	1.56	1.45	0.000 0	2.541 9	-0.158 7	-0.651 8	20 694 9	16.573 8
37	4,6-二氯-1,2-二硝基苯	2.42	2.25	2.17	0.000 0	2.277 1	-0.649 4	-1.6103	41 301 2	9.570 4
38	3,5-二硝基苯甲醇	0.53	0.86	1.25	0.000 0	3.486 6	0.228 1	-0.120 4	38 534 4	0.000 0
39	3,4-二硝基苯甲醇	1.09	0.91	1.31	0.000 0	3.613 8	-0.054 9	-0.271 3	38 852 7	0.000 0
40	2,4,6-三氯-1,3-二硝基苯*	2.19	2.42	2.51	0.000 0	1.018 2	-1.623 1	-1.539 6	42 958 8	13.852 5
41	2,3,5,6-四氯-1,4-二硝基苯	2.74	2.68	2.55	0.000 0	0.000 0	-1.220 3	0.159 5	36 322 0	23.183 5
42	2,4,5-三氯-1,3-二硝基苯*	2.59	2.60	2.62	0.000 0	1.064 5	-1.550 5	-1.539 6	42 958 8	13.733 7

* 化合物为检验集化合物;

1.2 电拓扑状态指数

电拓扑状态指数的计算方法详见文献 [4-8], 其计算步骤简述如下:

(1) 非氢原子间的拓扑距离 (d_{ij}): 以目标分子中非氢原子间的最短化学键数目为分子拓扑中非氢原子间的拓扑距离.

(2) 计算原子固有状态值 I

$$I = \{ (2N)^2 \cdot \delta^v + 1 \} / \delta, \quad (1)$$

其中 δ 和 δ^v 分别定义为: $\delta = \sigma - h$, $\delta^v = \sigma + \pi + n - h$

上式中 σ 为某非氢原子在分子图中与其他相邻非氢原子形成的 σ 键轨道中的电子数; h 为键合在该非氢原子上的氢原子数; v 是该非氢原子价电子层的电子数; π 和 n 分别表示某非氢原子与其他非氢原子键合形成的 π 键电子数以及未成键的孤对电子数.

(3) 计算分子中其他非氢原子对固有状态值的影响:

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

(4) 分子中非氢原子的电拓扑状态指数 S_i 为:

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

对于常见有机物而言, 每一种化合物有 41 个 ETSI 分子结构描述子, 每个描述子所对应的分子结构基元见文献 [9].

1.3 结构活性相关分析

本文以 42 种硝基苯类化合物所包含的 41 种原子类型的电拓扑指数描述子作为自变量 X , 以硝基苯类化合物的半抑制生长浓度为因变量 Y 构建数据集, 应用最佳子集回归程序对变量进行选择, 建立 QSAR 预测模型. 模型质量以相关系数 r^2 、LOO 交叉验证相关系数 q_{LOO}^2 、对外部检验集的预测相关系数 r_p^2 及均方根偏差 RMS 评价.

2 结果与讨论

2.1 描述子对样本化合物的分辨率

42 种硝基苯化合物中只含有 C—C 键、C = C 共轭双键、C—N 键、C—C \ C—Br \ N = O 键等化学键, 所有化合物的 41 种 ETS 描述子中只有 10 种描述子 ($x_1, x_3, x_6, x_{1b}, x_{22}, x_{24}, x_{25}, x_{39}, x_{40}, x_{41}$) 具有非零数值, 而其中描述子 x_{41} 只有化合物 6-碘-1,3-二硝基苯不为 0, 显然该描述子没有明确的统计意义, 利用剩余 9 种描述子对化合物的相似性进行分析, 定义样本中任意 2 个化合物在 9 个描述子构成的 9 维空间的模式距离 D_{ij} 为:

$$D_{ij} = \sqrt{\sum_{r=1}^9 (x_{ir} - x_{jr})^2}, \quad (4)$$

式 (4) 中 r 为描述子编号, x_i 及 x_j 为硝基苯类化合物的电拓扑状态指数描述子, 模式距离 D_{ij} 值越小, 说明样本越相似, 若该值为 0 则表明结构描述子不能区别两者化合物. 通过计算, 所有样本间的最大距离为 916.6252 (1,2-二硝基苯与 2,3,4,5-四氯硝基苯), 最小距离为 0.0055 (3,4-二氯硝基苯与 3,5-二氯硝基苯), 从而说明 9 种电拓扑状态指数描述子对 42 种化合物的结构分辨率为 100%. 硝基苯类化合物的部分电拓扑状态指数描述子见表 1.

2.2 模型的建立与质量评价

将 42 种硝基苯化合物的活性按从小到大排列, 均匀抽取 30 种化合物作为训练集, 剩余 12 种化合物构建检验集. 设定变量间的自相关系数的最大值 r_{int} 为 0.70, 在不同数目变量条件下, 运行最佳子集回归程序, 对训练集样本描述子变量进行优化选择, 所得到的变量组合及相关统计量见图 1.

图 1 表明, 模型非交叉验证相关系数 r^2 为随着变量数的增加而增大, 相应的均方根偏差 RMSEE 逐渐减小, 而 LOO 预测相关系数呈抛物线变化, 在变量数为 5 时达到最大值, 与之相对应的预测均方根偏差 RMSEP 最小, 因此最佳模型应该是 5 个电拓扑指数描述子变量的组合, 计算表明这 5 个变量之间没有明显的自相关 (表 2). 应用 5 个描述子变量建立的 QSAR 模型方程见式 (5):

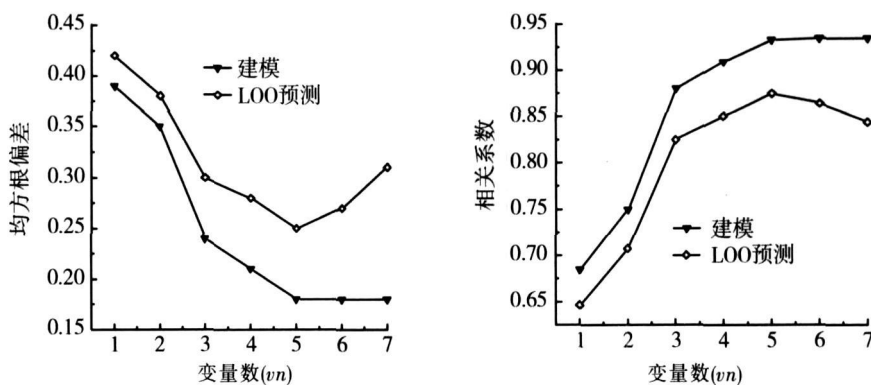


图 1 均方根误差及相关系数与变量数的关系

Fig.1 Various RMSE and r^2 with the number of variables

$$\log(IGC_{50}^{-1}) = - (1.10759 \pm 0.16568) + (0.12562 \pm 0.03600)x_1 - (0.28717 \pm 0.09925)x_{22} + (0.05005 \pm 0.00564)x_{25} + (0.08693 \pm 0.00666)x_{39} + (0.25282 \pm 0.02961)x_{40} \quad (5)$$

$$n = 30 \quad m = 5 \quad r^2 = 0.9327 \quad r = 0.9658 \quad RMSEE = 0.18 \quad F = 66.570 \text{ (Modeling)}$$

$$q^2 = 0.8749 \quad q = 0.9354 \quad RMSEP = 0.25 \text{ (LOO prediction)}$$

表 2 最佳模型描述子变量之间的相关性

Table 2 Correlation coefficients between various descriptors

No	x_1	x_{22}	x_{25}	x_{39}	x_{40}
x_1	1.0000				
x_{22}	0.2772	1.0000			
x_{25}	-0.2617	0.6503	1.0000		
x_{39}	-0.3814	-0.0020	-0.1010	1.0000	
x_{40}	0.1236	0.1335	-0.1764	-0.3193	1.0000

上式结果初步表明模型具有较好的拟合和预测能力, 其中 5 种描述子变量分别与 $-\text{CH}_3$ 、 $\geq \text{N}=\text{O}$ 、 $-\text{Cl}$ 及 $-\text{Br}$ 关联, 说明物质中这些结构单元的存在对硝基苯类化合物毒性有明显的影响, 这与其他 QSAR 方法的研究结果一致. 应用此模型对检验集中 12 种硝基苯化合物的生长毒性进行预测, 其预测相关系数 $r_p^2 = 0.8722$ 均方根误差为 0.26 结果表明模型具有较强的预测能力与稳健度, 模型对训练集与检验集的预测值与实验值的相关性见图 2

Cronin 等^[3] 根据硝基苯类化合物的毒性作用基本过程, 应用硝基化合物的正辛醇/水分配系数 ($\log K_{ow}$)、最低未占有分子轨道能量 (E_{LUMO}) 及最大受体超域性 (A_{max}) 参数对 42 种化合物建立了两种预测模型:

$$\log(IGC_{50}^{-1}) = 0.206(\log K_{ow}) - 16.0(A_{max}) - 5.04 \quad (6)$$

$$n = 42 \quad r^2 = 0.897 \quad s = 0.229 \quad F = 180 \quad q^2 = 0.888$$

$$\log(IGC_{50}^{-1}) = 0.467(\log K_{ow}) - 1.60(E_{LUMO}) - 2.55 \quad (7)$$

$$n = 42 \quad r^2 = 0.881 \quad s = 0.246 \quad F = 154 \quad q^2 = 0.866$$

上述模型方程中 A_{max} 与 E_{LUMO} 反映了化合物的亲电反应能力, $\log K_{ow}$ 表征化合物的疏水性质, 基于该 3 种参数的响应曲面分析方法对 42 种硝基苯类化合物的毒性具有较好的预测能力, 但文献中没有报道模型对外部数据集的预测结果. 与此相比, 本文应用电拓扑指数所建立的模型具有更高的估计相关系数, 同样具有良好的预测能力.

3 结论

本文研究了部分硝基苯类化合物对四膜虫生长的抑制毒性, 利用电拓扑指数表征化合物的分子结构并建立相应的 QSAR 模型, 结果表明影响该类化合物毒性的主要结构基团为 $-\text{CH}_3$ 、 $\geq \text{N}=\text{O}$ 、 $-\text{Cl}$ 与 $-\text{Br}$ 所建立模型不仅具有较好的拟合能力, 同时具有较强的预测能力.

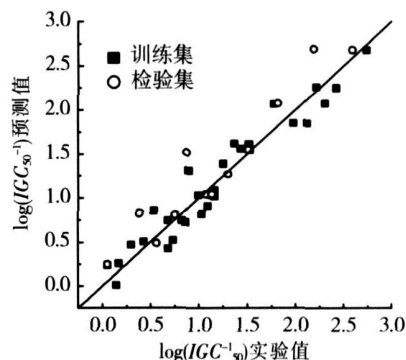


图 2 硝基苯的实验 $pIGC_{50}$ 值与预测值(ESTI)

Fig.2 Plot of $\log(IGC_{50}^{-1})$ values observed vs predicted

- [8] Choon Sup Raa, Seok Chan K in. π -Bridging 1, 3-Heterocyclic aromatic rings and the first hyperpolarizability of nonlinear optical compounds[J]. J Mol Struct(Theochim), 2004, 677(1/3): 173-178
- [9] Janczarek I, Fuks Luc J, Sahraoui B, et al. Third-order nonlinear optical figure of merits for conjugated TTF-quinone molecules[J]. J Phys Chem B, 2005, 109(20): 10 179-10 183
- [10] Nandi Prasanta Kumar, Mandal K. Theoretical study of static second-order nonlinear optical properties of push-pull heteroquinonoid dimers[J]. J Mol Struct(Theochim), 2006, 760(1/3): 235-244
- [11] Brasselet S, Cherioux E, Audebert P, et al. New octupolar star-shaped structures for quadratic nonlinear optics[J]. Chem Mater, 1999, 11(7): 1915-1920
- [12] Shu Y C, Gong Z H, Shu C F. Synthesis and characterization of nonlinear optical chromophores with conformationally locked polyenes possessing enhanced thermal stability[J]. Chem Mater, 1999, 11(6): 1628-1632
- [13] Galli S, Masciocchi N, Cariati E, et al. Quest for second-harmonic-generation-active coordination polymers: synthesis and properties of silver(I) pyrimidinates[J]. Chem Mater, 2005, 17(19): 4815-4824
- [14] Zhu X L, You X Z, Zhang Y. A novel approach to calculation of the second-order nonlinear optical susceptibilities of organic crystals based on energy-band theory[J]. Chem Phys, 2000, 254(2/3): 287-296
- [15] Li Zhong'an, Dong Shoucheng, Yu Gui, et al. Novel second-order nonlinear optical main-chain polyurethanes: adjustable subtle structure, improved thermal stability and enhanced nonlinear optical property[J]. Polymer, 2007, 48(19): 5520-5529
- [16] Ugurlu G, Kasap E, Kantarci Z, et al. A theoretical study of the linear, nonlinear optical properties and conformational analysis of 3-phenylthiophene and its fluoro derivatives with torsional dependence[J]. J Mol Struct, 2007, 834-836, 508-515
- [17] 王娟娟, 朱小蕾, 陈凯. 1, 2, 3, 4-四氢喹啉衍生物的第一超极化率的理论研究[J]. 南京师大学报: 自然科学版, 2006, 29(2): 66-72
- [18] Exner K, Schleyer P von R. Planar hexacoordinate carbon: a viable possibility[J]. Science, 2000, 290: 1937-1940

[责任编辑: 顾晓天]

(上接第 71 页)

[参考文献]

- [1] 徐镜波, 景体淞. 鲤鱼组织 ATPase 活性抑制和构效分析[J]. 高等学校化学学报, 1998, 19(12): 1920-1924
- [2] 徐满, 张爱茜, 韩朔暎, 等. 取代硝基苯类化合物的 3D-QSAR 研究[J]. 科学通报, 2001, 46(19): 1614-1617
- [3] Cronin M T D, Gregory B W, Schultz T W. Quantitative structure-activity analyses of nitrobenzene toxicity to tetrahymena pyrifomis[J]. Chem Res Toxicol, 1998, 11(8): 902-908
- [4] Kier L B, Hall L H. An electrotopological state index for atoms in molecules[J]. Pharmaceutical Research, 1990, 7(8): 801-807
- [5] Hall L H, Mohny B K, Kier L B. The electrotopological state: Structure information at the atomic level from molecular graphs[J]. J Chem Inf Comput Sci, 1991, 31(1): 76-82
- [6] Hall L H, Mohny B K, Kier L B. The electrotopological state: An atom index for QSAR[J]. Quant Struct-Act Relat, 1991, 10(1): 43-51
- [7] Kier L B, Hall L H. An index of electrotopological state of atoms in molecules[J]. J Math Chem, 1991, 7(1): 229-241
- [8] Hall L H, Kier L B, Brown B B. Molecular similarity based on novel atom type electrotopological state indices[J]. J Chem Inf Comput Sci, 1995, 35(6): 1074-1080
- [9] 王连生. 有机污染化学[M]. 北京: 高等教育出版社, 2004, 437-438

[责任编辑: 顾晓天]