

QSPR 方法预测乙腈溶剂中有机化合物的酸性解离常数

谢 昆, 乔 澍, 付 川, 程 聪

(重庆三峡学院化学与环境工程学院, 重庆 404100)

摘要:应用电拓扑状态指数为描述符,建立了乙腈溶剂中 43 个有机化合物的 pK_a 数值的定量构效关系(QSPR)模型。“逐一剔除”交叉验证的结果证明模型具有良好的稳定性和较强的预测能力. 建立的最佳七元回归模型为

$$pK_a = 0.225E_2 - 0.146E_4 - 1.929E_5 - 7.799E_6 - 5.049E_8 + 0.5327E_9 - 1.553E_{10} + 20.528,$$
该模型的相关系数 $R=0.986$, 标准偏差 $s=0.532$, 且无随机相关. 研究结果表明电拓扑状态指数能够有效地预测有机化合物的 pK_a 值.

关键词:电拓扑状态指数; QSPR; pK_a 值; 有机化合物; 乙腈

中图分类号: O621.16 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2010.10.003

Prediction of pK_a of organic compounds in acetonitrile with QSPR method

XIE Kun, QIAO Shu, FU Chuan, CHENG Cong

(College of Chemistry and Environmental Engineering, Chongqing Three Georges University, Chongqing 404100, China)

Abstract: The quantitative structure-property relationships (QSPR) model for pK_a values of 43 organic compounds in acetonitrile were established by using electrotopological state indices as descriptors. Leave-one-out cross validation was performed, which confirmed the good stability and powerful prediction of this model. A satisfactory relationship is expressed as

$$pK_a = 0.225E_2 - 0.146E_4 - 1.929E_5 - 7.799E_6 - 5.049E_8 + 0.5327E_9 - 1.553E_{10} + 20.528,$$
where $R=0.986$, $s=0.532$ without any random correlation. The result indicates that the electrotopological state indices can be effectively used for the prediction of pK_a values of organic compounds.

Key words: electrotopological state indices; QSPR; pK_a value; organic compounds; acetonitrile

0 引言

有机化合物结构与活性/性质定量相关(QSAR/QSPR)的研究最初作为定量药物设计的一个研究分支领域,是为了适应合理设计生物活性分子的需要而发展起来的. 近 20 多年来,由于计算机

技术的发展和运用,使 QSAR/QSPR 的研究日益成熟,其应用范围也正迅速扩大^[1],尤其在化学物质生态环境毒理风险预测方面有着较为广泛的应用^[2-5],或通过研究有毒化合物的毒性来指导有毒物处理的过程改进^[6]. 目前一种较为简便快捷的方法是利用分子拓扑指数与化合物的性质、活性参数建立

收稿日期:2010-06-12;修回日期:2010-10-14

基金项目:教育部“春晖计划”(Z2007-1-630010),国家水体污染控制与治理重大专项(2009ZX07104-003-02)和国家科技支撑计划重点项目(2008BAD98B04)资助.

作者简介:谢昆,女,1979年生,博士/副教授. 研究方向:物理有机化学. E-mail:xiekun_79@163.com

通讯作者:乔澍,博士/副教授. E-mail:qiaoshu_76@163.com

QSPR/QSAR 模型来对化合物的物理化学性质参数进行评估和预测^[7-10].

化合物给予或接受质子的能力大小,一般通过该化合物的酸碱性,即酸性解离常数 pK_a 的大小来衡量^[11]. 而 pK_a 数据本身又可以作为一个参数,对一些量子化学计算方面的研究起到重要的辅助作用^[12-18].

本文利用 Hall 等^[19] 提出的电拓扑状态指数 (electrotopological state indices) 作为参数,建立乙腈溶剂中 43 种有机化合物 pK_a 值与电拓扑状态指数相关的数学模型. 由于电拓扑状态指数包含分子的原子类型、成键价态和连接度等信息,故可以较好地表征分子结构,并揭示一些影响有机分子 pK_a 值的因素.

1 理论方法

1.1 电拓扑状态指数

电拓扑状态指数 (electrotopological state indices, 简称 E 状态指数, E_n) 是根据分子中每一个非氢原子在分子中的拓扑环境和成键电子信息进行计算得到的一组数值. E 状态指数由两部分组成: 一部分是原子的本征状态值, 其反映成键原子自身的价态及与相邻原子的成键信息; 另一部分是分子中其他非氢原子的扰动或影响. 将原子的本征值与扰动值相加, 便得到原子的 E 状态指数, 即

$$E_i = I_i + \Delta I_i$$

式中, E_i 表示成键原子 i 的电拓扑状态指数, I_i 为某原子 i 的固有状态值, ΔI_i 是其他非氢原子 j 引起的增量.

对于复杂的有机污染物或复杂药物分子, 可能含有多个原子类型相同 (固有状态值相同) 的非氢原子, 将相同类型原子的 E 状态指数相加, 即得到原子类型的电拓扑状态指数.

1.2 变量选择

本研究中, 采用了快速多模型 (leaps and bounds) 回归分析的方法^[20], 根据此种方法可以得到一组含有不同变量个数的最优方程. 采用剩余平方和 (RSS) 作为从相同数目的自变量组成的子集集合中, 识别出最优子集的判据. 即相同自变量个数子集的 RSS 越小, 则由这些变量回归所得到的数学模型最优. 该法将寻找最优数学模型问题转化成寻找具有最小剩余平方和值子集的问题, 减少了每一子集的运算量.

2 结果讨论

2.1 乙腈溶剂中 43 种有机化合物 QSPR 建模

首先, 在表 1 中列出在乙腈溶剂中所研究的 43 种化合物所包含的所有局部结构和它们所对应的电拓扑状态指数固有值.

表 1 乙腈溶剂中有机物局部结构对应的电拓扑状态指数固有值

Tab. 1 The intrinsic values of atomic types of organic molecules in acetonitrile

structure formal	I_i	E_n	structure formal	I_i	E_n
	2.000	E_1		1.333	E_{12}
	1.667	E_2		2.500	E_{13}
	7.000	E_3		4.111	E_{14}
	6.000	E_4		3.625	E_{15}
	2.000	E_5		1.667	E_{16}
	2.500	E_6		8.000	E_{17}
	4.000	E_7		6.000	E_{18}
	3.500	E_8		2.500	E_{19}
	1.5	E_9		8.000	E_{20}
	2.333	E_{10}		2.120	E_{21}
	2.750	E_{11}		1.250	E_{22}

分别计算选取 43 种化合物中所有非氢原子的电拓扑状态指数值, 并根据快速多模型回归分析的方法, 在上述的 22 个变量当中选取了 $E_2, E_4, E_5, E_6, E_8, E_9, E_{10}$ 7 个变量建立七元回归模型. 通过程序拟合进行计算, 得到最终的七元回归模型如下

$$pK_a = 0.225E_2 - 0.146E_4 - 1.929E_5 - 7.799E_6 - 5.049E_8 + 0.5327E_9 - 1.553E_{10} + 20.528$$

通过该模型公式计算这 43 种化合物的 pK_a 值, 与它们的电拓扑状态指数值结果一起列于表 2 中. 表 2 中的所有 pK_a 实验值取自文献^[21] 和^[22].

用实验 pK_a 值对计算 pK_a 值作图, 得到图 1, 并得到相关系数 $R=0.986$, 标准偏差 $s=0.532$.

为了避免随机相关, 需要对该模型进行稳定性

表 2 乙腈溶剂中 43 种化合物的 pK_a 值计算结果

Tab. 2 Molecular structure descriptors for 43 compounds in acetonitrile

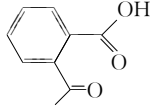
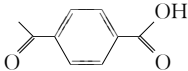
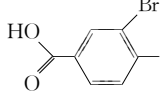
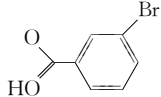
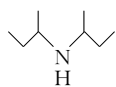
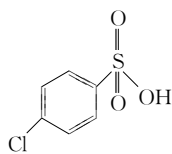
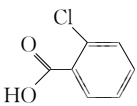
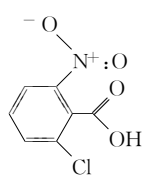
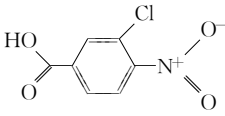
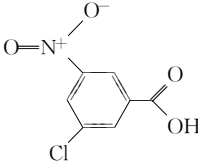
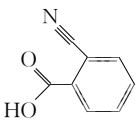
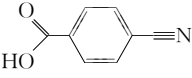
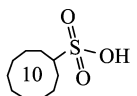
序号	structure formal	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
									Exp	Cal
1		-6.856	2.682	-0.651	0.000	0.000	0.000	0.000	20.51	19.9
2		-7.175	2.529	-0.561	0.000	0.000	0.000	0.000	19.95	19.6
3		-3.923	2.579	-0.088	0.000	0.000	0.000	0.000	20.03	19.4
4		-3.786	2.568	0.000	0.000	0.000	0.000	0.000	19.5	19.3
5		0.000	0.000	0.936	0.000	0.000	-0.436	0.000	18.81	18.5
6		-3.297	2.412	6.496	0.000	0.000	0.000	0.000	7.3	6.9
7		-5.504	2.481	0.000	0.000	0.000	0.000	0.000	19	18.9
8		-9.698	0.000	0.000	0.000	0.000	0.000	0.000	17.53	18.3
9		-8.463	0.000	0.000	0.000	0.000	0.000	0.000	18.28	18.6
10		-7.782	0.000	0.000	0.000	0.000	0.000	0.000	17.98	18.8
11		-5.714	2.657	0.000	0.000	0.000	0.000	0.000	19.07	18.9
12		-5.311	2.664	0.000	0.000	0.000	0.000	0.000	19.34	18.9
13		0.000	11.494	3.994	0.000	0.000	-4.879	0.000	8.02	8.5

表 2(续)

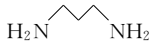

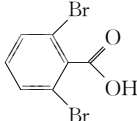
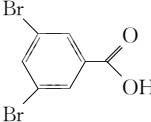
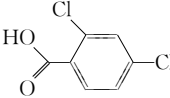
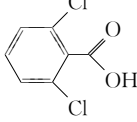
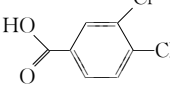
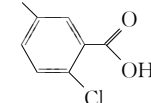
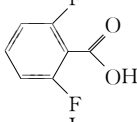
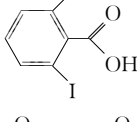
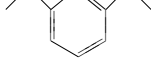
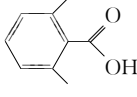
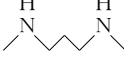

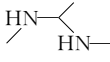
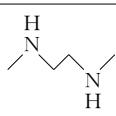



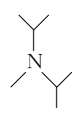
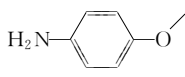

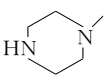
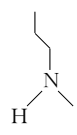

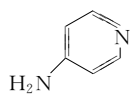

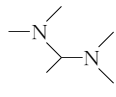
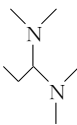
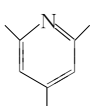
序号	structure formal	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
									Exp	Cal
14		0.000	0.000	0.000	0.000	0.000	-2.118	0.000	19.7	19.4
15		0.000	0.000	0.000	0.000	0.000	-3.480	0.000	19.1	18.7
16		-6.183	2.701	0.000	0.000	0.000	0.000	0.000	18.04	18.7
17		-5.711	2.703	0.000	0.000	0.000	0.000	0.000	18.86	18.8
18		-6.562	2.640	0.000	0.000	0.000	0.000	0.000	18.4	18.7
19		-7.507	2.652	0.000	0.000	0.000	0.000	0.000	18.16	18.5
20		-6.808	0.489	0.000	0.000	0.000	0.000	0.000	19.42	18.9
21		-6.991	2.630	0.000	0.000	0.000	0.000	0.000	18.75	18.6
22		-11.003	2.377	0.000	0.000	0.000	0.000	0.000	18.3	17.7
23		-5.551	2.849	0.000	0.000	0.000	0.000	0.000	18.27	18.9
24		-1.945	0.000	0.205	0.399	1.842	0.000	0.000	7.64	7.3
25		-5.454	2.576	-0.396	0.000	0.000	0.000	0.000	20.31	19.7
26		0.000	0.000	0.214	0.000	0.000	-0.638	0.000	20.39	19.8
27		0.000	0.000	0.245	0.000	0.000	-0.779	0.533	18.24	18.8
28		0.000	0.000	0.315	0.000	0.000	0.000	0.000	19.63	19.9

表 2(续)

序号	structure formal	E_2	E_4	E_5	E_6	E_7	E_8	E_9	pK_a	
									Exp	Cal
29		0.000	0.000	0.410	0.000	0.000	0.000	0.000	19.57	19.7
30		0.000	0.000	0.167	0.000	0.000	-0.655	0.488	18.3	19.1
31		0.000	0.000	-0.064	0.000	0.000	-1.626	0.453	19.57	19.1
32		-0.907	0.000	0.097	0.670	0.000	0.000	0.000	14.13	14.9
33		0.000	0.000	0.751	0.000	0.000	0.000	0.731	18.81	17.9
34		-2.152	0.000	0.370	0.000	1.386	0.000	0.000	11.86	12.3
35		0.000	0.000	0.051	0.000	0.000	-1.378	0.000	20.04	19.7
36		0.000	0.000	0.090	0.000	0.000	-1.136	0.706	18.07	18.7
37		0.000	0.000	0.268	0.000	0.000	-0.448	0.000	18.92	19.8
38		0.000	0.000	0.121	0.000	0.000	-1.378	0.000	20.04	19.6
39		-0.876	0.000	0.000	0.353	0.000	0.000	0.000	17.62	17.6
40		0.000	0.000	0.000	0.000	0.000	-0.362	0.000	19.56	20.3
41		0.000	0.000	-0.147	0.000	0.000	0.000	0.990	18.68	19.3
42		0.000	0.000	0.005	0.000	0.000	-0.393	1.143	19.27	18.5
43		-0.217	0.000	0.244	0.711	0.000	0.000	0.000	14.98	14.5

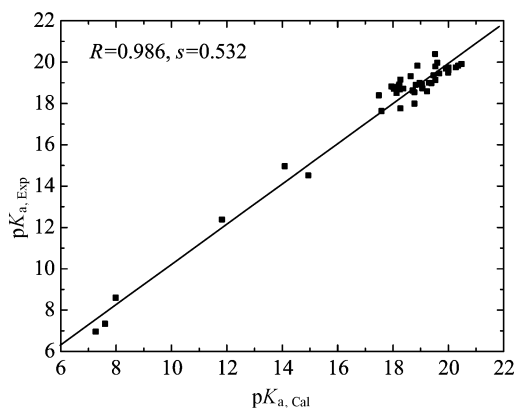


图 1 pK_a 值的计算值与实验值

Fig. 1 Predicted vs experimental pK_a

校检。由于研究的样本较少(43 个),所以采用“逐一剔除”的方法,即每次剔除一个化合物,剩余的 42 个化合物建模,求其相关系数。结果为小于 0.986 的有 4 个,在 0.986~0.987 之间的有 31 个,大于 0.987 的有 8 个。而平均值为 0.9864,与 0.986 也相当接近,所以该模型中不存在有显著影响的异常值,该模型中不存在随机相关,具有可以被接受的总体稳定性。

2.2 电拓扑状态指数的意义

从所建的 QSPR 模型中可以看出,乙腈溶剂中有机化合物的 pK_a 值更多地取决于静电作用。电拓扑状态指数中包含了较多分子中非氢原子的固有特性、成键价态信息以及拓扑环境等分子结构信息,其计算公式如下:

$$E_i = I_i + \Delta I_i$$

$$I_i = \frac{(2/N_i)^2 \cdot \delta_i^v + 1}{\delta_i}, \Delta I_i = \sum_{j \neq i} \frac{I_j - I_i}{(r_{ij})^2}$$

式中, I_i 为非氢原子的固有状态值; ΔI_i 为其他非氢原子对固有状态 I_i 的影响值; N_i 为原子 i 的主量子数; δ_i 定义为 $\delta_i = \sigma_i - h_i$, σ_i 为原子 i 在其 σ 成键轨道中电子数, h_i 为与原子 i 直接键合的氢原子数; 原子点价 δ_i^v 定义式为

$$\delta_i^v = \sigma_i + \pi_i + m_i - h_i$$

其中, π_i 和 m_i 分别为原子 i 与其他非氢原子键合所用的 π 键电子数以及未成键的孤对电子数。由 I_i 的定义式可知: ①主量子数 N 可以反映原子的大小; ② δ_i^v 可以用来表示非氢原子的成键情况, 其中 $\sigma_i + \pi_i$ 反映了该原子的价键信息; ③ δ_i 则呈现了各原子局部的拓扑环境。由 ΔI_i 的定义式可知, 它可以较好地反映分子中各非氢原子之间的电性相互作用和拓

扑相互作用^[23]。

3 结论

在乙腈溶剂中, 我们建立了如下七元回归模型:

$$pK_a = 0.225E_2 - 0.146E_4 - 1.929E_5 - 7.799E_6 - 5.049E_8 + 0.5327E_9 - 1.553E_{10} + 20.528$$

其相关系数为 $R=0.986$, 并经“逐一剔除”法检验无随机相关性, 具有可以被接受的总体稳定性。

从以上模型可以看出, 许多有机化合物在乙腈溶剂中的酸碱性 (pK_a 值) 与自身的电拓扑指数之间存在着定量关系。7 个变量比较全面地揭示了影响有机化合物在乙腈溶剂中的酸碱性的各种电子效应 (如诱导效应、共轭效应与场效应等), 因此, 通过电拓扑指数可以有效地预测有机物 pK_a 值。

参考文献 (References)

- [1] 王连生, 韩朔睽. 有机定量结构-活性相关[M]. 中国环境科学出版社, 1993:1.
- [2] Cheng Yan, Chen Huiming, Yu Wenlian, et al. Eco-environmental toxicity risk prediction for substances of very high concern with QSAR approach[J]. Research of Environmental Sciences, 2009, 22: 817-822. 程艳, 陈会明, 余文莲, 等. QSAR 技术对高关注化学物质生态环境毒理风险预测[J]. 环境科学研究, 2009, 22: 817-822.
- [3] Qin Liang, Luo Si, Gao Shumei, et al. The QSPR investigation of the degradation of chlorinated contamination by zero-valent iron[J]. Environmental Chemistry, 2009, 28: 400-403. 秦良, 罗斯, 高树梅, 等. 零价铁降解氯代有机污染物的 QSPR 研究[J]. 环境化学, 2009, 28: 400-403.
- [4] Yang Xushu, Wang Xiaodong, Zhang Yiming, et al. Holographic quantitative structure-activity relationship for prediction of the toxicity of polybrominated diphenyl ether congeners[J]. Scientia Sinica Chimica, 2010, 40: 86-94. 杨旭曙, 王晓栋, 张一鸣, 等. 应用分子全息 QSAR 预测多溴二苯醚(PBDEs)的毒性[J]. 中国科学: 化学, 2010, 40: 86-94.
- [5] Feng Changjun, Mu Lailong, Yang Weihua, et al. A mathematical model between bioconcentration factors and topological indices of organic pollutants[J]. Acta Phys-Chim Sin, 2008, 24: 1 053-1 057. 冯长君, 沐来龙, 杨伟华, 等. 有机污染物的生物富集因子与拓扑指数的数学模型[J]. 物理化学学报, 2008, 24: 1 053-1 057.

- [6] Li Jilai, Hang Yechao, Geng Caiyun, et al. QSAR study of acute toxicities of phenylsulfonyl carboxylate compounds [J]. Chemical Journal of Chinese Universities, 2007, 28: 117-120.
李吉来, 杭烨超, 耿彩云, 等. 苯磺基羧酸酯类急性毒性的 QSAR 研究[J]. 高等学校化学学报, 2007, 28: 117-120.
- [7] Ren Biye, Xu You, Chen Guobin. A novel topological index for QSPR/QSAR study of organic compounds [J]. Acta Chimica Sinica, 1999, 57: 563-571.
任碧野, 许友, 陈国斌. 一个新的拓扑指数用于有机化合物的 QSPR/QSAR 研究[J]. 化学学报, 1999, 57: 563-571.
- [8] Cao Chenzhong. Distance-edge topological index-research on structure-property relationship of alkanes [J]. Acta Chimica Sinica, 1996, 54: 533-538.
曹晨忠. 距离-边数拓扑指数-饱和链烃类化合物的结构/性质相关性研究[J]. 化学学报, 1996, 54: 533-538.
- [9] Cao Chenzhong, Yuan Hua. Study on the cis/trans-isomerism of alkenes by topological approach[J]. Acta Phy-Chim Sin, 2005, 21: 360-366.
曹晨忠, 袁华. 烯烷顺反异构的拓扑方法研究[J]. 物理化学学报, 2005, 21: 360-366.
- [10] Li Mingjian, Feng Changjun. QSAR study on the regulator activity of substituted benzoic acids to plant growth using topological methods [J]. Journal of Harbin Institute of Technology, 2009, 41: 195-197.
李鸣建, 冯长君. 取代苯甲酸对植物生长调节活性的拓扑 QSAR 研究[J]. 哈尔滨工业大学学报, 2009, 41: 195-197.
- [11] Fu Yao, Liu Lei, Li Ruiqiong, et al. First-principle predictions of absolute pK_a 's of organic acids in dimethyl sulfoxide solution[J]. J Am Chem Soc, 2004, 126: 814-822.
- [12] Fu Yao, Liu Lei, Yu Haizhu, et al. Quantum-chemical predictions of absolute standard redox potentials of diverse organic molecules and free radicals in acetonitrile[J]. J Am Chem Soc, 2005, 127: 7 227-7 234.
- [13] Fu Yao, Liu Lei, Wang Yimin, et al. Quantum-chemical predictions of redox potentials of organic anions in dimethyl sulfoxide and reevaluation of bond dissociation enthalpies measured by the electrochemical methods[J]. J Phys Chem A, 2006, 110: 5 874-5 886.
- [14] Li Jianing, Liu Lei, Fu Yao, et al. What are the pK_a values of organophosphorus compounds? [J]. Tetrahedron, 2006, 62: 4 453-44 62.
- [15] Qi Xiujian, Liu Lei, Fu Yao, et al. Ab initio calculations of pK_a values of transition-metal hydrides in acetonitrile[J]. Organometallics, 2006, 25: 5 879-5 886.
- [16] Li Jianing, Fu Yao, Liu Lei, et al. First-principle predictions of basicity of organic amines and phosphines in acetonitrile[J]. Tetrahedron, 2006, 62: 11 801-11 813.
- [17] Shen Kuang, Fu Yao, Li Jianing, et al. What are the pK_a values of C-H bonds in aromatic heterocyclic compounds in DMSO? [J]. Tetrahedron, 2007, 63: 1 568-1 576.
- [18] Xie Kun, Qiao Shu, Cheng Cong. QSPR studies on pK_a values of organic compounds in DMSO[J]. Acta Chimica Sinica, 2009, 67: 231-237.
谢昆, 乔澍, 程聪. DMSO 溶剂中有机化合物 pK_a 的定量构效关系研究[J]. 化学学报, 2009, 67: 231-237.
- [19] Hall L H, Kier L B. Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information[J]. J Chem Inf Comput Sci, 1995, 35: 1 039-1 045.
- [20] Shen Qi, Xu Lu, Li Hua. Studies on the method of leaps and bounds regression[J]. Chemical Journal of Chinese Universities, 1997, 18: 544-546.
申琦, 许禄, 李华. 快速多模型回归分析方法的研究[J]. 高等学校化学学报, 1997, 18: 544-546.
- [21] Jover J, Bosque, R, Sales J. Neural network based QSPR study for predicting pK_a of phenols in different solvents[J]. QSAR Comb Sci, 2007, 26: 385-397.
- [22] Sooväli L, Kaljurand I, Kütt A, et al. Uncertainty estimation in measurement of pK_a values in nonaqueous media: A case study on basicity scale in acetonitrile medium [J]. Analytica Chimica Acta, 2006, 566: 290-303.
- [23] Yang Weihua, Feng Changjun. The quantitative structure-property relationships research of pK_a values of substituted carboxylic acids[J]. Journal of Jishou University(Natural Science Edition), 2006, 27: 84-89.
杨伟华, 冯长君. 羧酸类化合物 pK_a 的定量构效关系[J]. 吉首大学学报(自然科学版), 2006, 27: 84-89.