

DOI: 10.7524/AJE.1673-5897.20191006001

廖立敏. 取代苯酚类化合物生物毒性的模拟预测[J]. 生态毒理学报, 2021, 16(2): 254-259

Liao L M. Simulation and prediction of biological toxicity of substituted phenolic compounds [J]. Asian Journal of Ecotoxicology, 2021, 16(2): 254-259 (in Chinese)

取代苯酚类化合物生物毒性的模拟预测

廖立敏

内江师范学院化学化工学院, 内江 641100

收稿日期: 2019-10-06 录用日期: 2019-10-25

摘要: 以化合物中的非氢原子及非氢原子之间的关系为结构描述符, 对部分取代苯酚类化合物结构进行了参数化表征。通过多元线性回归(MLR)和偏最小二乘回归(PLS)建立了化合物结构与毒性关系模型, 分析了影响化合物毒性的结构参数, 并将结果与文献进行了对比。2个模型的相关系数分别达0.984和0.988, 模型具有良好的拟合能力与预测能力; 化合物苯环上的取代基越多, 该化合物可能具有较大的毒性值; 本文取得的结果明显优于文献结果。本文对于环境中的有机污染物的结构与生物毒性关系研究具有一定的参考价值。

关键词: 取代苯酚; 生物毒性; 结构表征; 构效关系

文章编号: 1673-5897(2021)2-254-06 中图分类号: X171.5 文献标识码: A

Simulation and Prediction of Biological Toxicity of Substituted Phenolic Compounds

Liao Limin

College of Chemistry and Chemical Engineering, Neijiang Normal University, Neijiang 641100, China

Received 6 October 2019 accepted 25 October 2019

Abstract: The structure of some substituted phenolic compounds was characterized by the non-hydrogen atoms and the relationship between them in the compounds. The relationship between compound structure and toxicity was established by multiple linear regression (MLR) and partial least squares regression (PLS). The effects of structural parameters on the toxicity of the compounds were analyzed and the results were compared with the literature's results. The correlation coefficients of the two models are 0.984 and 0.988, respectively. The models have good fitting ability and predictive ability. The compound with more substituents on the benzene ring has the larger toxicity value. The results obtained in this paper are obviously better than the results of the literature. This paper has certain reference value for the study of the relationship between structure and biological toxicity of organic pollutants in the environment.

Keywords: substituted phenol; biological toxicity; structural characterization; structure-activity relationship

苯酚类化合物是重要的精细化学品或中间体, 在农药、医药、炸药和染料等方面有广泛应用。苯酚

基金项目: 四川省教育厅科研项目(18ZB0323); 内江师范学院校级重点项目(18ZA04)

第一作者: 廖立敏(1981—), 男, 副教授, 研究方向为化合物分子结构与性质的关系, E-mail: liaolimin523@126.com

* 通讯作者 (Corresponding author), E-mail: liaolimin523@126.com

类化合物大多具有毒性,个别还具有较强的致癌作用。随着现代工农业的发展,越来越多的苯酚类化合物随着废水排放到环境中,给动植物造成严重的危害。因此,全面评估苯酚类化合物的毒性,对于规范其生产、管理、使用和排放具有重要的意义。定量构效关系(QSAR)利用计算机模拟预测有机污染物性质参数,简便、省时、节约实验资源且降低费用,在全面获取有机污染物各种参数方面显示出了明显的优势^[1-4]。QSAR研究中,化合物结构参数化是关键步骤之一,目前应用广泛的主要有二维结构表征法^[5-6]和三维结构表征法^[7-8]。二维结构表征法的优势是计算简单、快速,但无法分辨顺反异构、光学异构等现象;三维结构表征法的优势是基于化合物三维立体结构计算,对各种异构体能很好地区分,但计算量大、复杂难懂。本文基于化合物分子二维结构

构建一种简易的结构描述符对化合物结构参数化表达,通过多元线性回归(MLR)和偏最小二乘回归(PLS)进行建模,分析影响化合物毒性的结构因素,为苯酚类化合物结构-性质研究提供参考。

1 材料与方法(Materials and methods)

1.1 实验材料

以25个取代苯酚类化合物为研究样本,化合物毒性以其诱发浮萍萎黄的活性值的负对数(pC)表示,其实验值取自文献[9],结果如表1所示。 pC 值越大表示化合物对浮萍的毒性越强,反之越弱。

1.2 实验方法

化合物结构参数化是建立化合物结构与毒性关系模型的关键一环,认为化合物中的非氢原子及非氢原子之间的关系对化合物的生物毒性有影响,而

表1 化合物及其毒性活性值的负对数(pC)
Table 1 Compounds and their negative logarithm of toxic activity (pC)

序号 No.	化合物 Compounds	毒性值 (Exp.) Toxicity value (Exp.)	计算值1 (Cal.1) Calculated value 1 (Cal.1)	误差1 (Err.1) Error 1 (Err.1)	计算值2 (Cal.2) Calculated value 2 (Cal.2)	误差2 (Err.2) Error 2 (Err.2)
1	苯酚 Phenol	1.80	1.72	-0.08	1.72	-0.08
2	4-氯苯酚 4-chlorophenol	2.70	2.86	0.16	2.95	0.25
3	2,4-二氯苯酚 2,4-dichlorophenol	3.40	3.68	0.28	3.66	0.26
4	2,4,6-三氯苯酚 2,4,6-trichlorophenol	4.50	4.48	-0.02	4.42	-0.08
5	2,4,5-三氯苯酚 2,4,5-trichlorophenol	5.10	4.65	-0.45	4.71	-0.39
6	2,3,4,6-四氯苯酚 2,3,4,6-tetrachlorophenol	5.60	5.34	-0.26	5.35	-0.25
7	五氯苯酚 Pentachlorophenol	6.20	6.17	-0.03	6.33	0.13
8	4-氯-2-甲基苯酚 4-chloro-2-methylphenol	3.20	3.09	-0.11	3.24	0.04
9	2-甲基-6-氯苯酚 2-methyl-6-chlorophenol	2.40	2.79	0.39	2.67	0.27
10	4-氯-3-甲基苯酚 4-chloro-3-methylphenol	3.20	3.25	0.05	3.25	0.05
11	2,6-二氯-4-甲基苯酚 2,6-dichloro-4-methylphenol	3.40	3.93	0.53	3.70	0.30
12	2,4,6-三氯-3-甲基苯酚 2,4,6-trichloro-3-methylphenol	4.70	4.72	0.02	4.52	-0.18
13	2,4,5,6-四氯-3-甲基苯酚 2,4,5,6-tetrachloro-3-methylphenol	5.40	5.55	0.15	5.49	0.09
14	2-甲基苯酚 2-methylphenol	2.20	1.97	-0.23	1.98	-0.22
15	2,6-二甲苯酚 2,6-dimethylphenol	2.40	2.20	-0.20	2.26	-0.14
16	2,4-二甲苯酚 2,4-dimethylphenol	2.60	2.54	-0.06	2.54	-0.06
17	2,5-二甲苯酚 2,5-dimethylphenol	2.60	2.52	-0.08	2.49	-0.11
18	3,5-二甲苯酚 3,5-dimethylphenol	2.70	2.80	0.10	2.73	0.03
19	2,4,6-三甲苯酚 2,4,6-trimethylphenol	2.80	2.75	-0.05	2.84	0.04
20	2,3,5-三甲苯酚 2,3,5-trimethylphenol	2.80	2.86	0.06	2.80	0.00
21	3-甲基-5-乙基苯酚 3-methyl-5-ethylphenol	3.10	3.10	0.00	2.97	-0.13
22	4-氯-3,5-二甲苯酚 4-chloro-3,5-dimethylphenol	3.60	3.62	0.02	3.56	-0.04
23	4-氯-2,5-二甲苯酚 4-chloro-2,5-dimethylphenol	3.60	3.49	-0.11	3.55	-0.05
24	4-氯-2,6-二甲苯酚 4-chloro-2,6-dimethylphenol	3.40	3.31	-0.09	3.55	0.15
25	4-氯-3-甲基-5-乙基苯酚 4-chloro-3-methyl-5-ethylphenol	4.00	4.01	0.01	4.13	0.13

氢原子仅仅影响与其直接相连的非氢原子的取值。非氢原子的取值除与其直接相连的氢原子有关外,还与非氢原子自身的电子结构相关,首先在参考文献[10-11]的基础上,利用式(1)对化合物中的非氢原子进行赋值。

$$Z_i = [(n_i - h_i) / (m_i - n_i - 1)]^{1/2} \quad (1)$$

式中: i 为原子在分子中的编码, n_i 为非氢原子 i 的价电子数, m_i 为原子核外电子总数, h_i 为与其直接连接的氢原子数。

不同类型的非氢原子对化合物的毒性值影响不尽相同,而相同类型的非氢原子对化合物的毒性值影响具有加和性,因而需要将化合物中的非氢原子进行分类。参阅文献[12-14]方法将化合物中的非氢原子分为4类,与 k ($k \leq 4$)个非氢原子直接相连的非氢原子为第 k 类非氢原子,如与2个非氢原子直接相连的仲碳原子为第2类非氢原子,以此类推。不同类型的非氢原子自身对化合物生物毒性的影响按式(2)进行累加。

$$x_k = \sum_{i \in k} Z_i \quad (k=1, 2, 3, 4) \quad (2)$$

式中: k 表示非氢原子 i 的原子类型, Z_i 按式(1)计算。根据非氢原子的分类,对于一个有机化合物分子中最多含有4类非氢原子,因此最终可得到4个非氢原子自身对化合物生物毒性的影响项,用 x_1 、 x_2 、 x_3 和 x_4 表示。

不同类型的非氢原子的关系对化合物的生物毒性的影响可能不同,相同类型的非氢原子之间的关系对化合物毒性的影响具有加和性。非氢原子之间的关系并非某种具体的相互作用,这种关系随着非氢原子之间的距离的增加而减弱,随着非氢原子的某种性质的增加而加强,式(3)可以满足这一要求。

$$x_r = m_{rl} = \sum_{i \in n, j \in l} Z_i \times Z_j \times \exp(-\alpha \times d_{ij}^2) \quad (n=1, 2, 3, 4; n \leq l \leq 4) \quad (3)$$

式中: Z 按式(1)计算; d_{ij} 为非氢原子 i 、 j 之间的相对距离(即键长之和与碳碳单键键长的比值,如果 i 、 j 之间有多条路径,则以最短的为准); n 和 l 为原子所属类型, $\alpha=0.5$ 。4类非氢原子之间可以有10种不同的组合,即 m_{11} 、 m_{12} 、 \dots 、 m_{44} ,简写为 x_5 、 x_6 、 \dots 、 x_{14} , m_{13} 表示第1类非氢原子与第3类非氢原子之间的关系,以此类推。这样一来,对于一个化合物经参数化表达后最多可得14个变量(结构描述符),暂将其称为化合物非氢原子及其关系(non-hydrogen atoms and their relationship, NATR)。

运用多元线性回归(MLR)及偏最小二乘回归

(PLS)进行建模,“留一法”对模型进行交互检验。

2 结果与讨论 (Results and discussion)

将化合物进行结构参数化表征后得到14个变量,但由于研究样本中不具有第4类非氢原子,因而与第4类非氢原子相关的 x_4 、 x_8 、 x_{11} 、 x_{13} 、 x_{14} 这5个变量全为“0”,其余9个非全“0”变量用于建模分析。

变量数达到9个,而样本数仅有25个,在建模之前有必要对变量进行筛选。剔除与化合物生物毒性相关性不大的变量,寻找最佳变量组合进行建模。运用逐步回归对变量进行筛选,依据变量的显著性大小依次将变量引入模型。逐步回归中建模相关系数(r)及标准偏差(SD)的变化情况如图1和图2所示。

由图1可知,建模的 r 随着变量的引入逐渐增大,起初增大的趋势非常明显,当变量数达到3个时,建模的 r 接近最大值,此后 r 略有增大,但增大趋势不明显。同样,由图2可知,SD随着变量的引入逐渐变小,起初变小的趋势非常明显,当变量数达到3个时,SD接近最小值,此后SD略减小,但减小趋势不明显。综上,应该选择3个变量进行建模,此时符合样本数(N)/变量数(n) ≥ 5 的经验规则,三变量模型如式(4)计算。

$$pC = 6.385 - 3.853 \times x_1 + 2.401 \times x_3 - 0.293 \times x_7 \quad (4)$$

建模: $N=25$, $r=0.984$, $SD=0.216$, $F=219.822$;交互检验: $N=25$, $r_{cv}=0.979$, $SD_{cv}=0.248$, $F_{cv}=165.810$ 。

建模的 r 高达0.984,说明模型具有良好的拟合能力;“留一法”交互检验相关系数(r_{cv})高达0.979,并且比建模的 r 略小,说明模型具有良好的预测能力与稳定性。判断模型质量好坏的另一个指标是

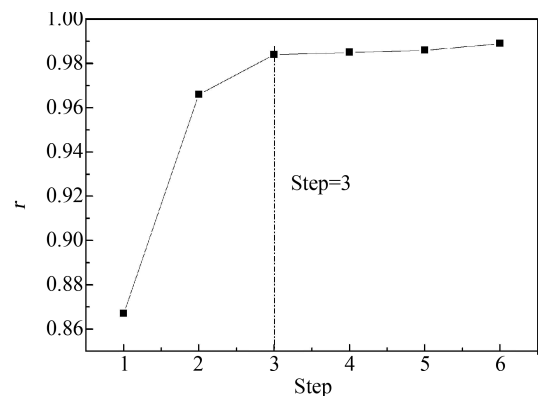


图1 相关系数(r)在逐步回归中的变化情况

Fig. 1 Changes of correlation coefficient (r) in stepwise regression

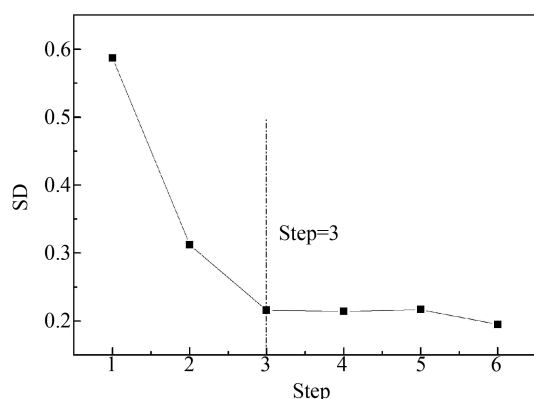


图2 标准偏差(SD)在逐步回归中的变化情况

Fig. 2 Changes of standard deviation (SD) in stepwise regression

SD,优良的模型通常要求SD与数值范围之比处于10%范围内。本文所建模型的SD为0.216,而化合物毒性值最大值-最小值=6.20-1.80=4.40,0.216/4.40=4.91%,远<10%,说明模型对化合物的生物毒性值预测误差在可接受的范围内。

为进一步分析影响化合物生物毒性的结构因素,采用偏最小二乘回归(PLS)建模和分析。将化合物结构描述符作为自变量 X ,化合物诱发浮萍萎黄的活性参数 pC 作为因变量 Y 建立PLS模型(M2)。PLS模型(M2)的主成分数(A)为6, r 为0.988,SD为0.174, r_{CV} 为0.980。从 r 、 r_{CV} 及SD值来看,模型(M2)也具有较好的拟合效果、预测能力和稳定性,并且建模效果还优于模型(M1),可以用于预测取代苯酚类化合物的毒性值和分析影响化合物毒性的结构因素。

25个样本在PLS前2个主成分得分空间散点分布如图3所示,由图3可知,所有的样本得分点都落在95%置信度的椭圆置信圈内,说明本文构建的化合物结构描述符能恰当表达该类化合物的结构特征,并且在PLS模型中得到反映。

变量重要性投影(VIP)可以反映各个变量与化合物的生物毒性相关性大小,通常认为VIP值>1的变量与所研究的性质相关性大,而其他变量与所研究的性质相关性可能较小。VIP如图4所示,由图4可知, x_5 、 x_3 和 x_2 的VIP值>1,说明这3个变量与化合物的毒性值相关性较大。进一步分析发现,VIP值最大的 x_5 为第1类非氢原子之间的关系对化合物毒性值的影响,第1类非氢原子对应的是化合物取代基末端原子; x_3 为第3类非氢原子自身对化合物毒性值的影响,第3类非氢原子对应的大部分是

苯环上被取代后的碳原子, x_5 和 x_3 这2个变量都与化合物的取代基数量有关。综上,化合物取代基越多,该化合物的生物毒性值可能就越大,这与表1中的数据特征基本相符。化合物产生毒性通常需要透过生物脂质膜,而化合物的体积与疏水性与其透过脂质膜的能力密切相关,当化合物分子都比较小时,化合物的疏水性起主导作用。本文研究样本中的取代基都为疏水性基团,取代基越多,该化合物的疏水性就越强,越易透过脂质膜而产生毒性。

2个模型对化合物的毒性值进行了预测,预测值分别为表1的Cal.1和Cal.2,Err.1和Err.2分别为其误差。模型预测值与实验值相关图如图5所示。由图5可知,所有的样本点都落在正方形的45°对角线上及附近,说明模型预测值与实验值高度相关,两者之间相异性小。2个模型对化合物毒性值预测误差的绝对值分布如图6所示。由图6可知,绝大部分样本的预测误差的绝对值都小于模型的2倍标

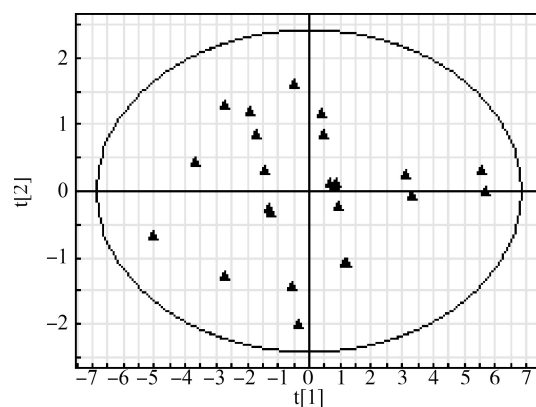


图3 样本在前2个主成分得分分布

Fig. 3 Distribution of the top 2 principal component scores of the sample

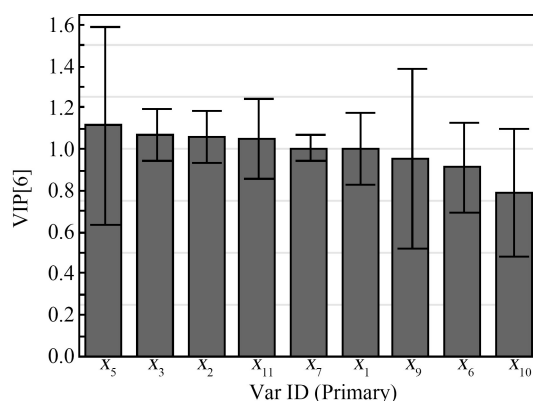


图4 变量重要性投影图

Fig. 4 Projection of variable importance

准偏差(2SD)。模型(M1)中,有2个样本的预测误差的绝对值超出了2SD的范围,而模型(M2)中所有样本的预测误差的绝对值均处于2SD范围内,说明模型(M2)对化合物毒性值的预测比模型(M1)更为准确,这与上述分析结果一致。

文献[15]采用DFT-B3LYP计算描述符,运用多元线性回归(MLR)进行了建模研究,将文献结果与本文结果进行了对比(表2)。

由表2可知,本文所建模型的 r 及 r_{CV} 明显大于文献结果,而SD明显小于文献结果,说明本文所建模型优于文献结果。另外,文献采用的分子结构表征方法DFT-B3LYP需经过分子结构优化,计算量大、复杂,而本文构建的结构描述符是基于分子二维

结构计算得到,无需结构优化,计算量小,简单易懂。

以化合物中的非氢原子及非氢原子之间的关系构建结构描述符,对部分取代苯酚类化合物结构进行了参数化表征。通过MLR和PLS方法建立了化合物结构与毒性关系模型,模型具有良好的拟合能力与预测能力。分析了影响化合物毒性的结构参数,并将结果与文献进行了对比,化合物苯环上的取代基越多,该化合物可能就有较大的毒性值,本文取得的结果明显优于文献结果。本文所构建的化合物结构描述符简单、易懂、计算方便,对于环境中的有机污染物的结构与生物毒性关系研究具有一定的参考价值。但结构描述符还存在不能区分顺反异构、光学异构等现象,这些在后续研究中都应予以考虑。

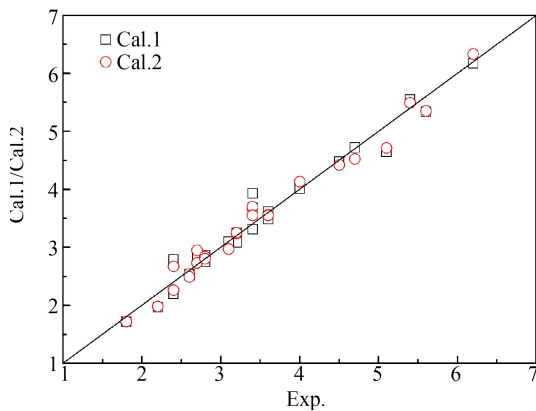


图5 模型预测值与实验值的相关图

Fig. 5 Correlation diagram between model predicted values and experimental values

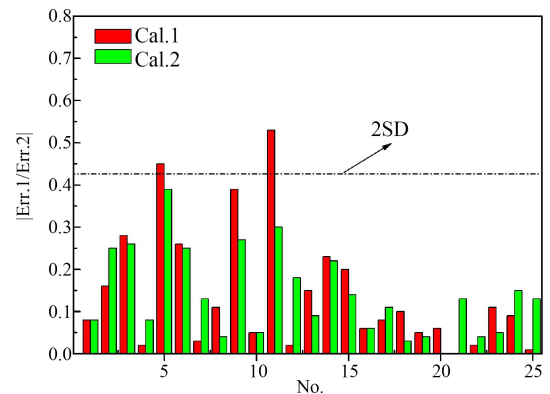


图6 模型对样本毒性的预测误差

注:SD表示标准偏差。

Fig. 6 Model prediction error of samples' toxicity

Note: SD stands for standard deviation.

表2 模型比较

Table 2 Model comparison

序号 No.	结构表征方法 Characterization method	建模方法 Modeling method	建模变量数/主成分数 Number of variables/ principal component	相关系数(r) Correlation coefficient (r)	交互检验相关系数(r_{CV}) Cross-check correlation coefficient (r_{CV})	标准偏差(SD) Standard deviation (SD)
1 ^[15]	DFT-B3LYP	MLR	2	0.972	0.965	0.283
2*	NATR	MLR	3	0.984	0.979	0.216
3*	NATR	PLS	6	0.988	0.980	0.174

注:*表示本文取得的结果;DFT-B3LYP表示密度泛函理论;ATR表示化合物非氢原子及其关系;MLR表示多元线性回归;PLS表示偏最小二乘回归。

Note: * means the results are obtained in this study; DFT-B3LYP is density functional theory; NATR is non-hydrogen atoms and their relationship; MLR is multiple linear regression; PLS is partial least squares regression.

参考文献(References):

[1] 舒茜. 取代芳烃与6种生物毒性的定量构效关系[J]. 四川师范大学学报:自然科学版, 2011, 34(3): 401-405
Shu Q. QSAR studing on the acute toxicities of substitu-

ted aromatic hydrocarbon to six kinds of organisms based on their molecular connectivity index [J]. Journal of Sichuan Normal University: Natural Science, 2011, 34(3): 401-405 (in Chinese)

- [2] Pasha F A, Srivastava H K, Srivastava A, et al. QSTR study of small organic molecules against *Tetrahymena pyriformis* [J]. *QSAR & Combinatorial Science*, 2010, 26(1): 69-84
- [3] Roy K, Ghosh G. QSTR with extended topochemical atom indices. 2. Fish toxicity of substituted benzenes [J]. *Journal of Chemical Information and Computer Sciences*, 2004, 44(2): 559-567
- [4] 杨伟华, 冯长君. 取代芳烃对蝌蚪的急性毒性的拓扑研究[J]. *徐州师范大学学报: 自然科学版*, 2005, 23(4): 53-56
Yang W H, Feng C J. The acute toxicity of substituted aromatics to tadpoles and QSAR study [J]. *Journal of Xuzhou Normal University: Natural Science Edition*, 2005, 23(4): 53-56 (in Chinese)
- [5] 邹丽云, 陈亚东, 尤启冬, 等. 2D-QSAR 和 HQSAR 研究 6-O-芳基酮内酯衍生物的定量构效关系[J]. *中国药科大学学报*, 2010, 41(3): 208-215
Zou L Y, Chen Y D, You Q D, et al. 2D-QSAR and HQSAR study on quantitative structure-activity relationship of 6-O-aryl ketolides derivatives [J]. *Journal of China Pharmaceutical University*, 2010, 41(3): 208-215 (in Chinese)
- [6] Xue C B, Ding Q, Luo W C, et al. 2D-QSAR studies of benzaldehyde, benzoic acid, benzaldehyde thiosemicarbazone, and their derivatives as insect phenoloxidase inhibitors [J]. *Chinese Journal of Pesticide Science*, 2008, 10(3): 275-281
- [7] 仝建波, 秦尚尚, 雷珊, 等. 4-羟氨基- α -吡喃酮甲酰胺类似物的 3D-QSAR 及分子对接研究[J]. *原子与分子物理学报*, 2019, 36(3): 397-405
Tong J B, Qin S S, Lei S, et al. 3D-QSAR and molecular docking of 4-hydroxyamino α -pyranone carboxamide analogues [J]. *Journal of Atomic and Molecular Physics*, 2019, 36(3): 397-405 (in Chinese)
- [8] 徐闻曦, 蔡晓然, 郑小娇, 等. TNNI₃K 抑制剂 3D-QSAR 的研究及虚拟筛选[J]. *武汉工程大学学报*, 2018, 40(5): 485-493
Xu W X, Cai X R, Zheng X J, et al. Virtual screening and 3D-QSAR study of TNNI₃K inhibitors [J]. *Journal of Wuhan Institute of Technology*, 2018, 40(5): 485-493 (in Chinese)
- [9] 王连生, 支正良. 分子连接性与分子结构——活性[M]. 北京: 中国环境科学出版社, 1992: 220-286
Wang L S, Zhi Z L. *Molecular connectivity and molecular structure—activity*[M]. Beijing: China Environmental Science Press, 1992: 220-286 (in Chinese)
- [10] 余训民, 余训爽. 分子价连接性指数中杂原子价点价计算新方法及应用[J]. *有机化学*, 2001, 21(9): 658-667
Yu X M, Yu X S. A new method for calculation valence delta of heteroatoms in molecular valence connectivity topological index and its application [J]. *Chinese Journal of Organic Chemistry*, 2001, 21(9): 658-667 (in Chinese)
- [11] 陈传品, 王鹏, 张国宇, 等. 分子连接性指数与羟基自由基反应活性定量关系[J]. *哈尔滨工业大学学报*, 2002, 34(4): 521-524, 528
Chen C P, Wang P, Zhang G Y, et al. Quantitative structure-activity relationships in hydroxyl radical reactions of organic chemicals [J]. *Journal of Harbin Institute of Technology*, 2002, 34(4): 521-524, 528 (in Chinese)
- [12] Liao L M, Huang X, Lei G D. Structural characterization and octanol/water partition coefficient (LogP) prediction for oxygen-containing organic compounds [J]. *Chinese Journal of Structural Chemistry*, 2017, 36(8): 1243-1250
- [13] 廖立敏, 李建凤, 雷光东. 饮用水中挥发性有机物结构与色谱保留时间的关系[J]. *环境化学*, 2017, 36(4): 710-715
Liao L M, Li J F, Lei G D. Relationship between structure and gas chromatography retention time of volatile organic compounds in drinking water [J]. *Environmental Chemistry*, 2017, 36(4): 710-715 (in Chinese)
- [14] 李建凤, 黄茜, 雷光东. 部分含氯芳烃结构与孔雀鱼半数致死浓度(-logLC₅₀)定量活性相关研究[J]. *生态毒理学报*, 2017, 12(3): 732-738
Li J F, Huang X, Lei G D. Study on the correlation between structure and quantitative activity of half lethal concentration(-logLC₅₀) to guppy for some chlorinated aromatic compounds [J]. *Asian Journal of Ecotoxicology*, 2017, 12(3): 732-738 (in Chinese)
- [15] 顾云兰, 朱琳, 费正皓, 等. 取代苯酚诱发浮萍萎黄活性预测研究[J]. *分子科学学报*, 2013, 29(3): 212-216
Gu Y L, Zhu L, Fei Z H, et al. Study on the prediction of chlorosis activity of duckweed induced by substituted phenols [J]. *Journal of Molecular Science*, 2013, 29(3): 212-216 (in Chinese) ◆