

DOI: 10.7524/AJE.1673-5897.20171221002

堵锡华, 王超. 醇和酚类污染物对欧洲林蛙蝌蚪及梨形四膜虫毒性的定量结构-活性模型[J]. 生态毒理学报, 2018, 13(6): 250-258

Du X H, Wang C. Quantitative structure-activity model of toxicity of alcohol and phenolic pollutants to *Rana temporaria* tadpoles and *Tetrahymena pyriformis* [J]. Asian Journal of Ecotoxicology, 2018, 13(6): 250-258 (in Chinese)

醇和酚类污染物对欧洲林蛙蝌蚪及梨形四膜虫毒性的定量结构-活性模型

堵锡华*, 王超

徐州工程学院 化学化工学院, 徐州 221018

收稿日期: 2017-12-21 录用日期: 2018-02-28

摘要: 醇和酚类等有机化合物作为重要的工业原料, 广泛应用于医药卫生、有机合成、食品工业等领域, 但生产中排放于环境的这些物质, 会对生物造成一定的毒性作用。为建立包含醇和酚类有机污染物对欧洲林蛙蝌蚪及梨形四膜虫毒性的定量结构-活性相关性模型, 计算了 227 种有机污染物的分子连接性指数和分子形状指数, 优化筛选了分子连接性指数的⁰X、¹X、²X、⁴X 和⁵X_c、分子形状指数的 K₁ 和 K₂ 共 7 种参数, 将这 7 种结构参数作为神经网络输入层变量, 110 种有机污染物对欧洲林蛙蝌蚪的毒性值作为输出层变量, 采用 7:8:1 的网络结构方式, 构建了令人满意的对欧洲林蛙蝌蚪毒性的神经网络预测模型, 方程总相关系数 r 为 0.988, 毒性预测值与实验值之间的平均误差为 0.14。为检验指数的普适性, 同样用这 7 个结构参数与 117 种醇和酚类化合物对梨形四膜虫的毒性进行分析, 所得神经网络模型的总相关系数达到 0.997, 对梨形四膜虫毒性的预测值与实验值之间的平均误差仅为 0.065, 结果表明, 所建模型具有良好的预测有机污染物对林蛙蝌蚪及梨形四膜虫急性毒性的能力。

关键词: 有机污染物; 欧洲林蛙; 梨形四膜虫; 分子连接性指数; 分子形状指数; 神经网络

文章编号: 1673-5897(2018)6-250-09 中图分类号: X171.5 文献标识码: A

Quantitative Structure-Activity Model of Toxicity of Alcohol and Phenolic Pollutants to *Rana temporaria* Tadpoles and *Tetrahymena pyriformis*

Du Xihua*, Wang Chao

School of Chemistry and Chemical Engineering, Xuzhou University of Technology, Xuzhou 221018, China

Received 21 December 2017 accepted 28 February 2018

Abstract: Organic compounds, such as alcohols and phenols, were important industrial raw materials. They were widely used in pharmaceutical industry, organic synthesis industry, food industry, etc. However, emission of these organic compounds had toxic effects on organisms in the environment. In order to establish quantitative structure-activity relationship model of the toxicity of organic contaminants to *Rana temporaria* tadpoles and *Tetrahymena pyriformis*, the relationship between molecular structure of 227 kinds of organic contaminants and the toxicity to *Rana temporaria* tadpoles and *Tetrahymena pyriformis* was analyzed. Moreover, molecular connectivity indices and molecular shape indices of organic compounds were calculated. The molecular connectivity indices ⁰X, ¹X, ²X, ⁴X

基金项目: 国家自然科学基金项目(No.21472071, No.21473081)

作者简介: 堵锡华(1963—), 男, 教授, 研究方向为环境污染构效学研究, E-mail: 12dxh@sina.com

and 5X_c , and molecular shape indices K_1 and K_2 , were selected. Then, the seven indices were used as input layer variables of neural network, the toxicity of 110 organic contaminants to *Rana temporaria* tadpoles was used as output layer variable and the 7:8:1 network structure was adopted to establish a satisfying neural network model. The total correlation coefficient r was 0.988. The mean error between the predicted value and experimental value was 0.14. In order to test universality, correlation between the structural parameters and the toxicity of 117 alcohol and phenolic compounds to *Tetrahymena pyriformis* was also analyzed by using the same method. The total correlation coefficient r was 0.997. The mean error between the predicted value and experimental value was 0.065. The results showed that the model had good predictive ability of the acute toxicity of organic contaminants to *Rana temporaria* tadpoles and *Tetrahymena pyriformis*.

Keywords: organic contaminant; *Rana temporaria*; *Tetrahymena pyriformis*; molecular connectivity index; molecular shape index; neural network

随着工业生产的快速发展,大量污染物被排放进入环境,导致水体中含有大量挥发性卤代烃、多环芳烃等有机污染物,这些污染物毒性大、累积性强、难以降解,有的在环境中能长期滞留,具有致癌、致畸、致突变性及对内分泌有干扰作用等^[1-2],对生态系统产生严重影响,危害人体健康,故世界多国将相当数量化合物列入优先控制环境污染物黑名单^[3-4]。酚类化合物作为重要的工业原料,广泛应用于石化、农药、医药和印染等行业,但它们具有“三致”作用^[5];醇类化合物则被广泛用于食品工业、工业溶剂和燃料添加剂,但它们的排放也对人身健康造成严重威胁^[6],故针对这些污染物的生物毒性研究,越来越受到研究人员的关注,有众多的研究成果见有报道^[7-10]。为了能快速获得这些污染化合物的生物毒性数据,近年来,利用定量结构-活性相关性(QSAR)研究方法来预测化合物的急性毒性,受到科研工作者的高度重视^[11-12],为了能提高预测环境污染物对生物毒性的准确度,本文利用在智能模式识别中广泛应用于化学^[13]、材料科学^[14]、药学^[15]、环境科学^[16]等领域的BP神经网络方法,对文献[17-19]所列的110种有机污染物对欧洲林蛙蝌蚪的急性毒性pIGC₅₀、117种醇和酚类化合物对梨形四膜虫的毒性pIGC₅₀,建立了分子连接性指数和分子形状指数与生物毒性之间的神经网络预测模型,所得毒性预测结果优于文献研究方法,本工作对研究有机污染物对生物的毒性效应以及环境的风险评价具有重要的现实意义。

1 研究方法(Research technique)

1.1 有机污染物毒性数据来源

110种包含醛、酮、酯、胺、醇、酚、卤代烃和芳香烃等有机污染物对欧洲林蛙蝌蚪的毒性数据 pIGC₅₀

(pIGC₅₀ 为对林蛙蝌蚪或梨形四膜虫的 50% 生长抑制浓度的负对数,单位为 mmol·L⁻¹)来源于文献[17-18],50 种酚类化合物对梨形四膜虫的毒性数据(pIGC₅₀)来源于文献[17,19],67 种醇类化合物对梨形四膜虫的毒性数据(pIGC₅₀)来源于文献[17,20],化合物及毒性实验值(Exp.)具体数据见表 1 和表 3。

1.2 分子连接性指数和形状指数的计算

用 Chemoffice Chem3D 9.0 绘图软件绘制 110 个有机污染物、50 个酚类化合物和 67 个醇类化合物的分子结构图,再在 MATLAB 软件中,应用文献方法自编的程序^[21],计算所有 227 个化合物分子的连接性指数和形状指数。首先对 110 个有机污染物的结构指数与其对欧洲林蛙蝌蚪的毒性进行最佳变量子集回归分析,通过优化筛选,发现选取分子连接性指数的⁰X、¹X、²X、⁴X 和⁵X_c(5 种指数分别代表 0 阶、1 阶、2 阶、4 阶路径指数和簇项指数)、分子形状指数的 K_1 和 K_2 (分别代表 1 阶、2 阶形状特征参数)共 7 种参数时,相关性最佳,相关数据列表 1。

2 模型的建构(The construction of the model)

2.1 回归分析

将 110 种有机污染物对欧洲林蛙蝌蚪毒性 pIGC₅₀,与优化筛选的 7 种分子结构参数,进行回归分析,得到方程:

$$\begin{aligned} \text{pIGC}_{50} = & -0.658 ^0X + 1.181 ^1X + 1.588 ^2X - 1.537 \\ & ^4X - 1.418 ^5X_c - 0.409 K_1 + 0.291 K_2 + 1.022 \end{aligned} \quad (1)$$

$$N=110, n=7, r=0.902, r^2=0.814, R_{\text{adj}}^2=0.801, S=0.519, F=63.846$$

式(1)中: N 、 n 、 r 、 r^2 、 R_{adj}^2 、 S 、 F 分别为样本数、变量数、相关系数、判定系数、调整的可决系数、标准误差、Fischer 检验值。

表1 化合物对欧洲林蛙蝌蚪毒性的预测

Table 1 Prediction of the toxicity of compounds to tadpoles of *Rana temporaria*

序号 No	化合物 Compound	pIGC ₅₀		序号 No	化合物 Compound	pIGC ₅₀	
		Exp.	Pre.			Exp.	Pre.
1	乙二醇 Ethylene glycol	0.19	0.07	56	氯乙烷 Chloroethane	2.35	2.17
2*	甲醇 Methanol	0.24	0.22	57*	丁酸乙酯 Ethylbutanoate	2.37	2.26
3	乙醇 Ethanol	0.54	0.70	58	苯酞 Phthalide	2.37	2.65
4	丙酮 Acetone	0.54	0.68	59	2-羟基苯甲酰胺 2-Hydroxybenzamide	2.48	2.95
5	氨基甲酸甲酯 Methylcarbamate	0.57	0.61	60	香兰素 Vanillin	2.48	2.58
6	碳酰胺 Carbamide	0.60	0.61	61	N-异丁基-氨基甲酸乙酯 n-Isobutylurethane	2.50	2.52
7*	琥珀酰胺 Succinamide	0.70	0.63	62*	对乙氧基乙酰苯胺 p-Ethoxyacetanilide	2.55	2.61
8	甲基乙酰脲 Methyl acetyl urea	0.76	0.95	63	戊烷 Pentane	2.55	2.65
9	3-氯-1,2-丙二醇 3-Chloropropan-1,2-diol	0.77	0.86	64	2-甲氧基酚 2-Methoxyphenol	2.57	2.35
10	乙酰胺 Acetamide	0.77	0.68	65	溴乙烷 Bromoethane	2.57	2.36
11	频哪醇 Pinacol	0.81	0.84	66	1,2-二氯乙烷 1,2-Dichloroethane	2.63	2.19
12*	2-丙醇 Propan-2-ol	0.89	0.98	67*	2-甲基-2-丁烯 2-Methylbut-2-ene	2.64	2.42
13	2-甲基-2-丙醇 2-Methylpropan-2-ol	0.89	0.85	68	苯 Benzene	2.68	2.39
14	正丙醇 Propan-1-ol	0.96	0.98	69	乙酸戊酯 n-Pentyl acetate	2.72	2.98
15	丁酮 Butanone	1.04	1.19	70	戊酸乙酯 Ethyl pentanoate	2.72	2.56
16	乙酸甲酯 Methyl acetate	1.10	1.21	71	喹啉 Quinoline	2.72	2.90
17*	甲酸乙酯 Ethylformate	1.16	1.13	72*	间甲苯酚 m-Cresol	2.75	3.03
18	酒石酸乙二酯 Diethyl tartrate	1.22	1.17	73	对甲苯酚 p-Cresol	2.75	2.63
19	2-甲基-2-丁醇 2-Methylbutan-2-ol	1.24	1.22	74	吗啡 Morphine	2.76	2.79
20	戊酰胺 Pentanamide	1.30	1.85	75	胡椒醛 Piperonal	2.78	2.68
21	2-甲基丙醇 2-Methylpropan-1-ol	1.35	1.49	76	甲基苯基醚 Methylphenylether	2.82	2.80
22*	正丁醇 Butan-1-ol	1.42	1.37	77*	氯仿 Trichloromethane	2.85	2.95
23	氨基甲酸乙酯 Ethyl urethane	1.46	1.50	78	N,N-二甲基苯胺 N,N-dimethylaniline	2.85	2.68
24	N-乙基-氨基甲酸乙酯 N-ethylurethane	1.46	1.86	79	樟脑 Camphor	2.88	2.87
25	二乙醚 Diethylether	1.47	1.19	80	邻甲苯酚 o-Cresol	2.92	2.80
26	乙酸乙酯 Ethyl acetate	1.52	1.06	81	N-异戊基-氨基甲酸乙酯 n-Isopentylurethane	2.93	2.92
27*	3-戊酮 Pentan-3-one	1.54	1.44	82*	碘乙烷 Iodoethane	2.96	3.01
28	甲基叔丁基醚 Methyl t-butylether	1.55	1.43	83	苯乙酮 Acetophenone	3.04	2.96
29	三聚乙醛 Paraldehyde	1.60	1.64	84	1,4-二甲氧基苯 1,4-Dimethoxybenzene	3.05	3.13
30	吡啶 Pyridine	1.60	1.93	85	四氯化碳 Tetrachloromethane	3.14	3.15
31	3-甲基丁醇 3-Methylbutan-1-ol	1.64	1.69	86	N-苯基-氨基甲酸乙酯 n-Phenylurethane	3.22	3.22
32*	间苯二酚 Resorcinol	1.64	2.00	87*	香豆素 Coumarin	3.24	3.25
33	三醋酸甘油酯 Triacetin	1.64	1.63	88	1,3-二甲氧基苯 1,3-Dimethoxybenzene	3.35	3.32
34	2-戊酮 Pentan-2-one	1.72	1.75	89	1-辛醇 Octan-1-ol	3.40	3.63
35	乙酰乙酸乙酯 Ethyl acetoacetate	1.72	1.77	90	1-庚醇 Heptan-1-ol	3.42	3.30
36	安替吡啉 Antipyrine	1.89	1.93	91	间二甲苯 m-Xylene	3.42	3.08
37*	1,3-二氯-2-丙醇 1,3-Dichloropropan-2-ol	1.95	1.70	92*	2-丙基吡啶 2-Propylpyridine	3.48	3.56
38	乙酸丙酯 n-Propyl acetate	1.96	1.92	93	尼古丁 Nicotine	3.51	3.59
39	丙酸乙酯 Ethylpropanoate	1.96	2.08	94	戊酸丁酯 n-Butylpentanoate	3.60	3.45
40	苯胺 Aniline	1.96	2.30	95	丁香油酚 Eugenol	3.91	3.84
41	1,1-二乙氧基乙烷 1,1-Diethoxyethane	1.98	2.14	96	薄荷醇 Menthol	3.97	3.89
42*	柠檬酸三乙酯 Triethyl citrate	2.04	2.05	97*	6-溴-1,2-癸二醇 6-Bromo-1,2-decanediol	4.01	4.33
43	乙基硫醇 Ethylthiol	2.09	2.22	98	1,12-十二烷二醇 Dodecan-1,12-diol	4.02	4.26
44	对甲氧基乙酰苯胺 p-Methoxyacetanilide	2.09	2.21	99	萘 Naphthalene	4.19	3.94
45	邻苯二酚 Catechol	2.12	1.87	100	3-溴-1,2-癸二醇 3-Bromo-1,2-decanediol	4.25	4.32
46	对苯二酚 Hydroquinone	2.12	1.76	101	2-异丙基-5-甲基苯酚 2-Isopropyl-5-methylphenol	4.26	4.27
47*	正戊醇 Pentan-1-ol	2.15	2.06	102*	马钱子碱 Strychnine	4.34	4.38
48	N-丙基-氨基甲酸乙酯 n-Propylurethane	2.18	2.04	103	二苯胺 Diphenylamine	4.43	4.43
49	水合三氯乙醛 Chloral hydrate	2.22	2.20	104	2-溴-1,2-癸二醇 2-Bromo-1,2-decanediol	4.47	4.30
50	异戊酸乙酯 Ethylisobutanoate	2.24	2.41	105	对叔戊基苯酚 4-t-Pentylphenol	4.52	4.31
51	乙酸异丁酯 Isobutyl acetate	2.24	2.22	106	5-溴-1,2-癸二醇 5-Bromo-1,2-decanediol	4.54	4.33
52*	苯酚 Phenol	2.28	2.05	107*	1-壬醇 Nonan-1-ol	4.60	4.34
53	乙酸正丁酯 n-Butyl acetate	2.30	2.23	108	1-癸醇 Decan-1-ol	5.00	5.24
54	乙酰苯胺 Acetanilide	2.31	2.23	109	菲 Phenanthrene	5.25	5.16
55	苯基脲 Phenylurea	2.34	2.32	110	十一醇 Undecan-1-ol	5.30	5.01

注:pIGC₅₀ 表示毒性值;Exp.表示实验值;Pre.表示预测值。Note: pIGC₅₀ stands for toxicity value; Exp. stands for experiment value; Pre. stands for predicted value.

2.2 稳健性与“离域值”检验

为检验模型的稳健性及是否存在离域数据,用Jackknifed 逐一剔除法对模型检验,在 110 个分子中,依次剔除 1 个,用余下 109 个有机物分子进行回归,这样可建立 110 个方程,有 110 个相关系数 r 值(见表 2),这些相关系数基本维持在 0.9 左右,对这些 Jackknifed 相关系数 r 作雷达图(如图 1),以 0.89 为圆心,0.005 为间距,110 个分子的 r 值全部落在 0.890~0.915 之间,波动很小,说明不存在异常的毒性数据,另采用 MINITAB14 软件,用逐一剔除法检验得到的交互检验相关系数 R^2_{cv} 为 $0.771 > 0.5$,调整的可决系数 $R^2_{adj}(0.801)$ 与交互检验相关系数 $R^2_{cv}(0.771)$ 的差值为 0.03,小于 0.3,说明没有过拟合现象,模型有良好的稳健性。

2.3 神经网络模型的建构

为提高准确预测有机污染物对欧洲林蛙蝌蚪的毒性,用 BP 神经网络法进一步研究,将筛选得到的 7 种结构参数作 BP 法的输入层变量,对欧洲林蛙蝌蚪毒性 $pIGC_{50}$ 作输出层变量,综合 Andrea 及许禄等^[22]学者对隐含层变量选择定义的规则,隐含层按照 $2.2 > N/M \geq 1.4$,其中 $M=(I+1)H+(H+1)Q$,式中 N 为总样本数, M 为神经网络总权重, I, H, Q 分别为神经网络中输入层、隐含层和输出层的变量数。经

过计算,隐含层变量 H 可取 6、7 或 8,经运算比较,当 H 取 8 时所得模型的相关性最佳,故网络结构采用 7 : 8 : 1 方式。运算时,将 110 个分子样本分为训练集(每 5 个分子作为一组,取第 1、3、4 个分子,依次类推)、测试集(每 5 个分子的第 2 个)和验证集(每 5 个分子中的第 5 个),所得模型的总相关系数 r

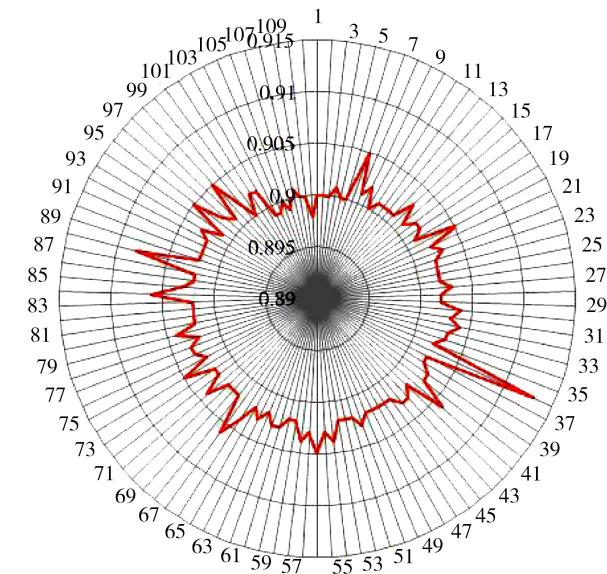


图 1 Jackknifed 相关系数 r 的雷达图

Fig. 1 Radar map of correlation coefficient r

表 2 Jackknifed 相关系数 r 的检验

Table 2 Inspection of Jackknifed correlation coefficient r

Remove molecule	r	Remove molecule	r	Remove molecule	r	Remove molecule	r	Remove molecule	r	Remove molecule	r
1	0.900	17	0.902	33	0.903	49	0.902	65	0.902	81	0.902
2	0.900	18	0.902	34	0.903	50	0.903	66	0.904	82	0.902
3	0.900	19	0.901	35	0.902	51	0.902	67	0.906	83	0.902
4	0.901	20	0.905	36	0.913	52	0.902	68	0.902	84	0.906
5	0.900	21	0.902	37	0.902	53	0.902	69	0.902	85	0.903
6	0.900	22	0.903	38	0.902	54	0.904	70	0.902	86	0.902
7	0.905	23	0.902	39	0.903	55	0.903	71	0.904	87	0.902
8	0.901	24	0.902	40	0.903	56	0.905	72	0.902	88	0.908
9	0.902	25	0.902	41	0.906	57	0.903	73	0.902	89	0.902
10	0.900	26	0.902	42	0.902	58	0.904	74	0.905	90	0.902
11	0.901	27	0.903	43	0.903	59	0.902	75	0.902	91	0.902
12	0.901	28	0.902	44	0.903	60	0.902	76	0.903	92	0.902
13	0.901	29	0.902	45	0.902	61	0.903	77	0.903	93	0.903
14	0.901	30	0.904	46	0.902	62	0.903	78	0.902	94	0.901
15	0.903	31	0.903	47	0.902	63	0.902	79	0.904	95	0.905
16	0.901	32	0.904	48	0.902	64	0.903	80	0.902	96	0.902

$=0.988$,训练集相关系数 $r_1=0.988$ 、测试集相关系数 $r_2=0.989$ 、验证集相关系数 $r_3=0.986$;利用该模型得到的有机污染物对欧洲林蛙蝌蚪毒性值 $pIGC_{50}$ 的预测值(Pre.)列于表1中(表1中序号有*号的分子为测试集分子),该预测值(Pre.)与实验值(Exp.)吻合度较好,两者的平均残差为0.14,它们的关系见图2。

3 指数普适性检验(Universal test)

为检验筛选出的7个分子连接性指数与形状指数的应用普适性,将文献所列的50个酚和67个醇共117个分子对梨形四膜虫的毒性进行分析,得到回归方程为:

$$pIGC_{50} = -0.446^0 X + 1.037^1 X + 1.381^2 X - 3.637^4 X - 0.204^5 X_c - 0.218 K_1 - 0.090 K_2 - 1.792 \quad (2)$$

$$N=117, n=7, r=0.924, r^2=0.853, R_{adj}^2=0.844, S=0.486, F=90.523$$

为检验模型的稳健性,用MINITAB软件对其进行交互检验,得到模型的 $r_{cv}^2=0.815$,而且 $R_{adj}^2(0.844)-r_{cv}^2(0.815)=0.029<0.3$,说明不存在“离域异常数据”,说明稳健性较好。对该模型用前面同样的分析方法和步骤,建立醇酚类化合物结构与毒性的神经网络模型,所得模型总相关系数为0.997,训练集相关系数 $r_1=0.997$ 、测试集相关系数 $r_2=0.997$ 、验证

集相关系数 $r_3=0.997$;这里3个集合的相关系数与总相关系数完全一致,而且均是超过0.99以上的高度相关性,利用该模型得到的醇酚类化合物对梨形四膜虫毒性 $pIGC_{50}$ 的预测值(Pre.)列于表3中(表中序号有*号的分子为测试集分子),该预测值(Pre.)与实验值(Exp.)吻合度也较为理想,两者平均残差为0.065,它们的关系见图3。

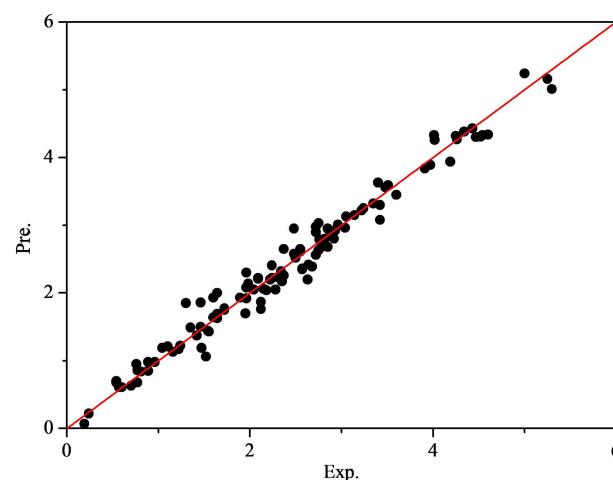


图2 $pIGC_{50}$ 的实验值与计算值关系图

Fig. 2 Relationship between experimental and calculated value of $pIGC_{50}$

表3 醇酚类化合物对梨形四膜虫毒性的预测

Table 3 Prediction of the toxicity of alcohol and phenolic compounds to *Tetrahymena pyriformis*

序号 No	化合物 Compound	$pIGC_{50}$		序号 No	化合物 Compound	$pIGC_{50}$	
		Exp.	Pre.			Exp.	Pre.
1	苯酚 Phenol	-0.431	-0.383	60	乙醇 Etanol	-1.991	-1.956
2*	对甲苯酚 p-Cresol	-0.192	-0.162	61	异丙醇 Isopropanol	-1.882	-1.817
3	间甲苯酚 m-Cresol	-0.062	-0.039	62*	丁醇 1-Butanol	-1.431	-1.605
4	2,5-二甲基苯酚 2,5-Dimethylphenol	0.009	0.324	63	2-丁醇 2-Butanol	-1.542	-1.686
5	3-氟苯酚 3-Fluorophenol	0.017	0.261	64	2-甲基-1-丙醇 2-Methyl-1-propanol	-1.372	-1.435
6	3,5-二甲基苯酚 3,5-Dimethylphenol	0.113	0.198	65	2-戊醇 2-Pentanol	-1.159	-1.030
7*	2,3-二甲基苯酚 2,3-Dimethylphenol	0.122	0.115	66	3-戊醇 3-Pentanol	-1.244	-1.290
8	3,4-二甲基苯酚 3,4-Dimethylphenol	0.122	0.124	67*	3-甲基-2-丁醇 3-Methyl-2-butanol	-0.996	-0.973
9	2,4-二甲基苯酚 2,4-Dimethylphenol	0.128	-0.014	68	2-甲基-1-丁醇 2-Methyl-1-butanol	-0.953	-1.170
10	2-乙基苯酚 2-Ethylphenol	0.176	0.187	69	3-甲基-1-丁醇 3-Methyl-1-butanol	-1.036	-0.956
11	2-氟苯酚 2-Fluorophenol	0.248	0.300	70	叔丁醇 tert-Butanol	-1.791	-1.772
12*	2-氯苯酚 2-Chlorophenol	0.277	0.007	71	己醇 Hexyl alcohol	-0.379	-0.412
13	3-乙基苯酚 3-Ethylphenol	0.299	0.289	72*	3,3-二甲基-1-丁醇 3,3-Dimethyl-1-butanol	-0.737	-0.737
14	2,6-二氯苯酚 2,6-Dichlorophenol	0.396	0.613	73	4-甲基-1-戊醇 4-Methyl-1-pentanol	-0.637	-0.720
15	3,4,5-三甲基苯酚 3,4,5-Trimethylphenol	0.418	0.421	74	庚醇 Heptan-1-ol	0.105	0.102
16	4-氟苯酚 4-Fluorophenol	0.473	0.318	75	1-辛烯-3-醇 1-Octen-3-ol	0.031	0.106
17*	4-异丙基苯酚 4-Isopropylphenol	0.473	0.558	76	1-壬醇 1-Nonanol	0.855	0.961
18	2-溴苯酚 2-Bromophenol	0.504	0.559	77*	2-壬醇 2-Nonanol	0.618	0.526
19	4-氯苯酚 4-Chlorophenol	0.545	0.558	78	3-乙基-2,2-二甲基-3-戊醇 3-Ethyl-2,2-dimethyl-3-pentanol	-0.169	-0.169

续表3

序号 No	化合物 Compound	pIGC ₅₀		序号 No	化合物 Compound	pIGC ₅₀	
		Exp.	Pre.			Exp.	Pre.
20	3-异丙基苯酚 3-Isopropylphenol	0.609	0.580	79	癸醇 Decyl alcohol	1.335	1.398
21	2-氯-5-甲基苯酚 2-Chloro-5-methylphenol	0.640	0.604	80	4-癸醇 4-Decanol	0.849	0.799
22*	4-溴苯酚 4-Bromophenol	0.681	0.651	81	四氢芳樟醇 Tetrahydrolinalool	0.340	0.343
23	4-氯-2-甲基苯酚 4-Chloro-2-methylphenol	0.700	0.862	82*	十一醇 1-Undecanol	1.954	1.807
24	3-叔丁基苯酚 3-Tert-butylphenol	0.730	0.728	83	十三醇 1-Tridecanol	2.449	2.480
25	4-氯-3-甲基苯酚 4-Chloro-3-methylphenol	0.795	0.767	84	2-甲基-3-丁烯-2-醇 2-Methyl-3-buten-2-ol	-1.389	-1.359
26	2-异丙基苯酚 2-Isopropylphenol	0.803	0.694	85	3-甲基炔醇 3-Methylbutynol	-1.311	-1.316
27*	3-氯-4-氟苯酚 3-Chloro-4-fluorophenol	0.842	0.777	86	反式-3-己烯-1-醇 Trans-3-hexen-1-ol	-0.777	-0.738
28	4-碘苯酚 4-Iodophenol	0.854	0.756	87*	5-己炔-1-醇 5-Hexyn-1-ol	-1.295	-1.085
29	4-叔丁基苯酚 4-Tert-butylphenol	0.913	0.914	88	3-甲基-1-戊炔-3-醇 3-Methyl-1-pentyn-3-ol	-1.322	-1.325
30	2,3,7-三甲基苯酚 2,3,7-Trimethylphenol	0.930	0.805	89	5-己烯-1-醇 5-Hexen-1-ol	-0.841	-0.852
31	2,4-二氯苯酚 2,4-Dichlorophenol	1.036	1.143	90	4-戊炔-2-醇 4-Pentyn-2-ol	-1.632	-1.771
32*	2-苯基苯酚 2-Phenylphenol	1.094	1.092	91	5-己炔-3-醇 5-Hexyn-3-ol	-1.404	-1.244
33	3-碘苯酚 3-Iodophenol	1.118	1.114	92*	3-庚炔-1-醇 3-Heptyn-1-ol	-0.323	-0.379
34	2,5-二氯苯酚 2,5-Dichlorophenol	1.128	1.114	93	4-庚炔-2-醇 4-Heptyn-2-ol	-0.616	-0.660
	4-氯-3,5-二甲基苯酚			94	3-辛炔-1-醇 3-Octyn-1-ol	0.017	0.036
35	4-Chloro-3,5-dimethylphenol	1.203	1.121	95	炔丙醇 Propargyl alcohol	-1.918	-1.957
	2-叔丁基-4,6-二甲基苯酚			96	2-丁烯醇 Crotonyl alcohol	-1.472	-1.486
36	2-(tert-Butyl)-4,6-dimethylphenol	1.245	1.258	97*	3-丁烯-2-醇 3-Buten-2-ol	-1.053	-1.054
37*	2,3-二氯苯酚 2,3-Dichlorophenol	1.271	1.169	98	2-丁烯-1,4-二醇 2-Butene-1,4-diol	-2.149	-2.121
	4-溴-6-氯-2-甲基苯酚			99	顺-2-戊烯醇 Cis-2-penten-1-ol	-1.105	-1.162
38	4-Bromo-6-chloro-2-methylphenol	1.277	1.437	100	3-戊烯-2-醇 3-Penten-2-ol	-1.401	-1.320
	4-溴-2,6-二甲基苯酚			101	反式-2-己烯-1-醇 Trans-2-hexen-1-ol	-0.472	-0.748
39	4-Bromo-2,6-dimethylphenol	1.278	1.094	102*	顺式-2-己烯-1-醇 Cis-2-hexen-1-ol	-0.777	-0.748
	2-叔丁基-4-甲基苯酚			103	反式-2-辛烯-1-醇 Trans-2-octen-1-ol	0.365	0.404
40	2-tert-Butyl-4-methylphenol	1.297	1.256	104	1,5-戊二醇 1,5-Pentanediol	-1.934	-1.866
41	2,4-二溴苯酚 2,4-Dibromophenol	1.403	1.568	105	DL-1,2-己二醇 DL-1,2-Hexanediol	-1.267	-1.246
42*	3,5-二氯苯酚 3,5-Dichlorophenol	1.562	1.439	106	1-丙醇 1-Propanol	-1.746	-1.643
43	2,4,6-三氯苯酚 2,4,6-Trichlorophenol	1.695	1.581	107*	2-甲基-2-丁醇 2-Methyl-2-butanol	-1.173	-1.202
44	4-溴-2,6-二氯苯酚 4-Bromo-2,6-dichlorophenol	1.779	1.777	108	新戊醇 Neopentyl alcohol	-0.870	-0.870
	2,6-二叔丁基-4-甲基苯酚			109	2,4-二甲基-3-戊醇 2,4-Dimethyl-3-pentanol	-0.705	-0.658
45	2,6-Di-tert-butyl-4-methylphenol	1.788	1.788	110	辛醇 Capryl alcohol	0.583	0.573
	4-氯-2-异丙基-5-甲基苯酚			111	DL-2-辛醇 DL-2-Octanol	0.001	0.096
46	4-Chloro-2-isopropyl-5-methylphenol	1.862	1.914	112*	十二烷醇 Dodecyl alcohol	2.161	2.166
47*	2,4,6-三溴苯酚 2,4,6-Tribromophenol	2.050	2.059	113	4-戊炔-1-醇 4-Pentyn-1-ol	-1.420	-1.469
48	2,4,5-三氯苯酚 2,4,5-Trichlorophenol	2.100	2.061	114	叶醇 Leaf alcohol	-0.809	-0.738
49	2,6-二苯基苯酚 2,6-Diphenylphenol	2.113	2.090	115	4-己烯-1-醇 4-Hexen-1-ol	-0.754	-0.795
	2,4-二溴-6-苯基苯酚			116	3-壬炔-1-醇 3-Nonyn-1-ol	0.340	0.365
50	2,4-Dibromo-6-phenylphenol	2.207	2.220	117*	1-己烯-3-醇 1-Hexen-3-ol	-0.811	-0.823
51	1,2-丁二醇 1,2-Butanediol	-2.048	-1.967				
52*	1,3-丁二醇 1,3-Butanediol	-2.301	-2.261				
53	1,4-丁二醇 1,4-Butanediol	-2.236	-2.215				
54	1,2-戊二醇 1,2-Pantanediol	-1.627	-1.544				
55	2-甲基-2,4-戊二醇 2-Methyl-2,4-pantanediol	-1.953	-1.975				
56	1,6-己二醇 1,6-Hexanediol	-1.495	-1.554				
57*	1,2-癸二醇 1,2-Decanediol	0.764	0.775				
58	1,10-癸二醇 1,10-Decanediol	0.224	0.216				
59	甲醇 Methanol	-2.666	-2.711				

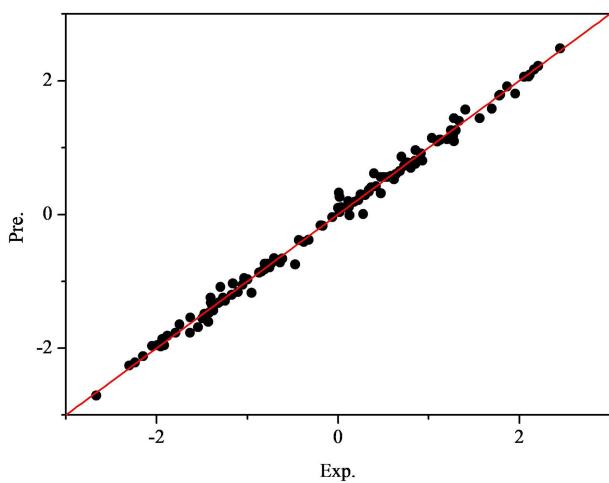
图 3 pIGC_{50} 的实验值与计算值关系图

Fig. 3 Relationship between experimental and calculated value of pIGC_{50}

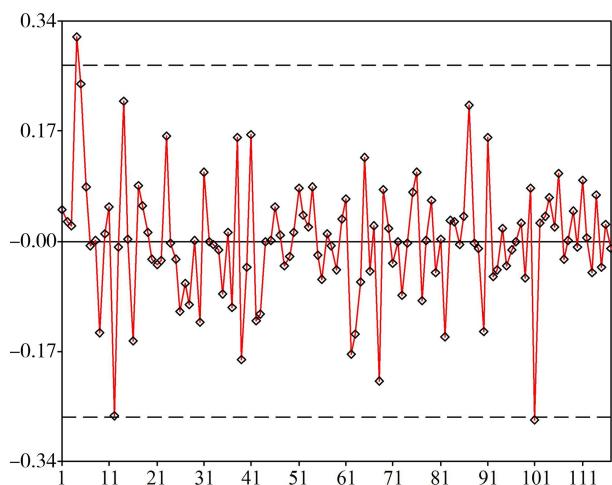


图 4 残差控制图

Fig. 4 Control chart of error

这里将预测残差作控制图(见图 4),可以看出,所有数据均在可控范围内,没有超出离域范围,只有对第 4 个分子 2,5-二甲基苯酚的预测残差偏大,这可能与苯环上连接的羟基和 2 个甲基之间的相关作用导致毒性值异常有关。

4 结果与讨论(Results and discussion)

通过计算 227 个有机污染物的分子连接性指数和形状指数,筛选了其中的 7 种指数与 110 种有机物分子对欧洲林蛙蝌蚪、117 种醇和酚类化合物对梨形四膜虫毒性的相关性分析,建立了对林蛙蝌蚪毒性预测的神经网络模型,预测毒性值与实验值较

为吻合,平均误差为 0.14,而且所选用的结构参数的普适性较强,它们与 117 种醇酚类化合物对梨形四膜虫的毒性之间,也能建立良好相关性的神经网络模型,总相关系数能达到 0.99 以上的高度相关,预测毒性值与实验值的平均误差仅为 0.065,通过逐一剔除法检验,可以看出,所建模型具有良好的稳定性和预测能力,并且具有普适应用性。从表 1 和表 3 化合物分子的结构与毒性大小关系可以看出,分子的毒性大小与分子的大小、所连接的基团性质、所处的位置和数量多少有密切的关系,化合物的毒性随着碳原子数的增加、文化度增大而增大,当碳原子数相同时,含有不饱和键越多,毒性会增大;如化合物的氢原子被卤原子取代,毒性也会增大,取代数量越多,毒性越强,毒性大小依次为:碘取代>溴取代>氯取代,这主要是卤素原子具有强烈的吸电子效应,使分子极性增加,易与酶系统结合,使毒性增大;芳烃的毒性效应则与苯环上的吸电子基团有关,如吸电子基团越多,毒性越强。分子连接性指数和分子形状指数正是蕴含了丰富的分子空间结构信息,能很好地反映毒性变化规律。

综上所述:(1)模型具有较好的预测能力。优化筛选分子连接性指数的 0X 、 1X 、 2X 、 4X 和 5X 、分子形状指数的 K_1 和 K_2 共 7 种分子结构参数,能充分反映分子的空间结构和电性结构信息,与化合物对生物的毒性有良好的相关性,用神经网络方法建构的模型预测的结果与毒性实验值吻合度较好。

(2)模型具有较强的普适应用型。筛选的分子结构参数不但与众多种类的有机污染物对林蛙蝌蚪毒性能较好相关,而且与醇类、酚类化合物对梨形四膜虫的毒性有良好的相关性,所建模型还具有良好的稳健性。

(3)神经网络模型的相关系数明显优于多元回归法模型,而且预测误差明显降低,说明分子结构参数与生物毒性之间具有良好的非线性关系。

参考文献(References):

- [1] 毕鹏禹,齐婧,吴昱,等.固相微萃取技术在环境污染物检测中的应用[J].理化检验-化学分册,2015,51(10):1487-1492
- Bi P Y, Qi J, Wu Y, et al. Application of solid phase microextraction in inspection of environmental pollutants [J]. Physical Testing and Chemical Analysis Part B: Chemical Analysis, 2015, 51(10): 1487-1492 (in Chinese)
- [2] Bastian K, Piotr M, Christine S. Quantifying the impact of

- immobile water regions on the fate of nitroaromatic compounds in dual-porosity media [J]. *Journal of Contaminant Hydrology*, 2016, 191: 44-53
- [3] 郭娜, 沈卫阳. 量子点在环境污染物检测中的应用研究进展[J]. 环境科学与技术, 2013, 36(6): 85-90
Guo N, Shen W Y. Progress of application of quantum dots in environmental pollutants analysis [J]. *Environmental Science & Technology*, 2013, 36(6): 85-90 (in Chinese)
- [4] 王斌, 余刚, 张祖麟, 等. 基于 QSAR 和 PCA 方法的硝基芳烃综合毒性评价[J]. 清华大学学报: 自然科学版, 2007, 47(3): 369-372, 376
Wang B, Yu G, Zhang Z L, et al. Evaluation of integrated toxicities of nitroaromatic compounds based on QSAR and PCA [J]. *Journal of Tsinghua University: Science and Technology*, 2007, 47(3): 369-372, 376 (in Chinese)
- [5] 顾云兰, 张淦杰. 酚类化合物对梨形四膜虫毒性的构效关系解析[J]. 化学研究与应用, 2016, 28(6): 810-817
Gu Y L, Zhang G J. Studies on QSAR for the toxicity of phenols against *Tetrahymena pyriformis* [J]. *Chemical Research and Application*, 2016, 28(6): 810-817 (in Chinese)
- [6] 郑景泉, 杨浩娜, 彭世文, 等. 醇类化合物对欧洲林蛙蝌蚪毒性的 QSAR 研究[J]. 山东农业科学, 2017, 49(2): 67-71
Zheng J Q, Yang H N, Peng S W, et al. QSAR study on toxicity of alcohol compounds to tadpole of European rana [J]. *Shandong Agricultural Sciences*, 2017, 49(2): 67-71 (in Chinese)
- [7] Toropov A A, Toropova A P, Marzo M, et al. QSAR models for predicting acute toxicity of pesticides in rainbow trout using the CORAL software and EFSA's Open-FoodTox database [J]. *Environmental Toxicology and Pharmacology*, 2017, 53: 158-163
- [8] Tugcu G, Ertürk M D, Saçan M T. On the aquatic toxicity of substituted phenols to *Chlorella vulgaris*: QSTR with an extended novel data set and interspecies models [J]. *Journal of Hazardous Materials*, 2017, 339: 122-130
- [9] 王婷, 林志芬, 田大勇, 等. 有机污染物的混合毒性 QSAR 模型及其机制研究进展[J]. 科学通报, 2015, 60(19): 1771-1780
Wang T, Lin Z F, Tian D Y, et al. QSAR models and their corresponding toxicity mechanism for mixture toxicity of organic pollutants [J]. *Chinese Science Bulletin*, 2015, 60(19): 1771-1780 (in Chinese)
- [10] Kar S, Roy K. QSAR modeling of toxicity of diverse organic chemicals to *Daphnia magna* using 2D and 3D descriptors [J]. *Journal of Hazardous Materials*, 2010, 177(1/2/3): 344-351
- [11] 赵永升, 黄莺, 赵继红, 等. 基于 QSAR 方法的离子液体毒性预测[J]. 化工学报, 2014, 65(5): 1616-1621
Zhao Y S, Huang Y, Zhao J H, et al. QSAR for predicting toxicity of imidazolium based ionic liquids [J]. *CIESC Journal*, 2014, 65(5): 1616-1621 (in Chinese)
- [12] 于洋, 王晓红, 闻洋, 等. 应用 Abraham 方程研究有机污染物对七种水生生物的毒性[J]. 环境化学, 2015, 34(1): 23-36
Yu Y, Wang X H, Wen Y, et al. Study on the toxicity of organic pollutants to seven aquatic organisms based on Abraham model [J]. *Environmental Chemistry*, 2015, 34(1): 23-36 (in Chinese)
- [13] 堵锡华. 用新的路径定位指数和神经网络研究多溴联苯醚理化性质[J]. 化工学报, 2014, 65(4): 1169-1178
Du X H. Physicochemical property of polybrominated diphenyl ethers by new path location index and neural network [J]. *CIESC Journal*, 2014, 65(4): 1169-1178 (in Chinese)
- [14] 堵锡华, 王超. 碳纳米管吸附能力的神经网络理论模型[J]. 环境化学, 2016, 35(7): 1445-1451
Du X H, Wang C. Research on the adsorption of carbon nanotubes to aromatic compounds by neural network [J]. *Environmental Chemistry*, 2016, 35(7): 1445-1451 (in Chinese)
- [15] Brahma S, Boztepe C, Kunkul A, et al. Modeling of drug release behavior of pH and temperature sensitive poly (NIPAAm-co-AAc) IPN hydrogels using response surface methodology and artificial neural networks [J]. *Materials Science and Engineering: C*, 2017, 75: 425-432
- [16] 李钦玲, 杨玉良. 人工神经网络预测取代芳烃类化合物对黑呆头鱼的毒性[J]. 计算机与应用化学, 2017, 34(5): 395-400
Li Q L, Yang Y L. Predicting the toxicity of substituted aromatic compounds to fathead minnows using Artificial Neural Network [J]. *Computers and Applied Chemistry*, 2017, 34(5): 395-400 (in Chinese)
- [17] 邓小龙, 陈渊, 谭泗桥, 等. 醇酚类化合物毒性的 QSAR 研究[J]. 环境科学学报, 2016, 36(12): 4490-4499
Deng X L, Chen Y, Tan S Q, et al. QSAR study on toxicities of alcohol and phenol compounds [J]. *Acta Scientiae Circumstantiae*, 2016, 36(12): 4490-4499 (in Chinese)
- [18] 苏强. 基于数据挖掘算法的环境毒物 QSAR 研究[D]. 上海: 上海大学, 2013: 83-86
Su Q. Research on QSAR of environmental toxicant based data mining [D]. Shanghai: Shanghai University, 2013: 83-86 (in Chinese)
- [19] 郭明, 许禄. 酚类化合物的 QSAR 研究[J]. 环境科学学报, 1998, 18(2): 122-126

- Guo M, Xu L. Quantitative structure-activity relationship for toxicity of phenols using regression analysis and neural network [J]. *Acta Scientiae Circumstantiae*, 1998, 18(2): 122-126 (in Chinese)
- [20] 王新颖, 张锦晖, 王丹丹, 等. 脂肪醇化合物对梨形四膜虫急性毒性的QSAR研究[J]. 计算机与应用化学, 2014, 31(6): 732-736
- Wang X Y, Zhang J H, Wang D D, et al. Prediction of acute toxicities of fatty alcohol to *Tetrahymena pyriformis* based on quantitative structure-activity relationship (QSAR) [J]. *Computers and Applied Chemistry*, 2014, 31(6): 732-736 (in Chinese)
- [21] 张婷, 梁逸曾, 赵晨曦, 等. 基于分子结构预测气相色谱程序升温保留指数[J]. *分析化学*, 2006, 34(11): 1607-1610
- Zhang T, Liang Y Z, Zhao C X, et al. Prediction of temperature-programmed retention indices from molecule structures [J]. *Chinese Journal of Analytical Chemistry*, 2006, 34(11): 1607-1610 (in Chinese)
- [22] 许禄, 邵学广. 化学计量学方法[M]. 北京: 科学出版社, 2004: 441
- Xu L, Shao X G. Methods of Chemometrics [M]. Beijing: Science Press, 2004: 441 (in Chinese) ◆