



DOI:10.13364/j.issn.1672-6510.20190243

数字出版日期: 2020-09-29; 数字出版网址: <http://kns.cnki.net/kcms/detail/12.1355.N.20200928.1325.001.html>

利用范数描述符预测苯衍生物对费氏弧菌 (*V. fischeri*) 的毒性效应

史琦玉¹, 贾青竹¹, 王强²

(1. 天津科技大学海洋与环境学院, 天津 300457; 2. 天津科技大学化工与材料学院, 天津 300457)

摘要: 苯衍生物属于水体环境典型污染物, 其水体生态毒性风险备受关注. 定量构效关系(QSAR)模型对研究有机物的环境风险评估具有重要作用. 本工作根据有机物的原子分布矩阵计算得到系列范数描述符, 据此建立 QSAR 模型, 研究 110 个苯衍生物对费氏弧菌(*V. fischeri*)的生态毒理效应. 研究结果表明: 本工作建立的 QSAR 模型能准确预测苯衍生物对 *V. fischeri* 的毒性, 其中, 模型统计结果相关性系数 $R^2 = 0.895$, 均方根误差为 0.241; 内部验证和外部验证表明了该模型的可靠性和稳定性; 应用域验证表明了本模型的广泛适用性. 由此推断, 本工作提出的范数描述符可能是有机物分子结构的一种根本描述, 据此建立的模型可以应用于有机污染物生态毒性评价和环境风险评价.

关键词: QSAR; 范数描述符; 苯衍生物; 毒性; 费氏弧菌(*V. fischeri*)

中图分类号: X131.2 文献标志码: A 文章编号: 1672-6510(2020)06-0029-08

Predicting Toxic Effects of Benzene Derivatives on *V. fischeri* with Norm Descriptors

SHI Qiyu¹, JIA Qingzhu¹, WANG Qiang²

(1. College of Marine and Environmental Sciences, Tianjin University of Science & Technology, Tianjin 300457, China;

2. College of Chemical Engineering and Materials Science, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Benzene derivatives are typical pollutants in water environment, and their ecological toxicity has attracted great attention. QSAR model plays an important role in assessing the environmental risk caused by organic compounds. In this project, a series of norm descriptors were calculated based on atomic distribution matrixes of organic compounds, and a QSAR model was established to study the ecotoxicological effects of 110 benzene derivatives on *Vibrio fischeri* (*V. fischeri*). The results showed that the newly established QSAR model could accurately predict the toxicity of benzene derivatives to *V. fischeri*. The correlation coefficient R^2 is 0.895 and the root-mean-square error is 0.241. Internal and external validation tests verified the reliability and stability of the model. Application domain analysis showed that this model was widely applicable. Accordingly, it can be concluded that the norm descriptors proposed in this research might be a fundamental description of the molecular structure of organic compounds, and the model could be applied to the ecotoxicity and environmental risk evaluations of organic pollutants.

Key words: QSAR; norm descriptors; benzene derivatives; toxicity; *V. fischeri*

苯衍生物包括苯胺、苯酚、硫酚、苯甲酸和抗生素(磺胺类与磺胺类增强剂)等, 在化工、医药和农业等领域广泛应用, 由此导致这些化合物在环境中显著检出; 同时, 此类化合物对水体环境产生生态毒性,

对人体健康存在潜在的风险, 因此, 有必要对此类化合物的生态风险进行评估^[1].

有机物对费氏弧菌(*V. fischeri*)的生态毒性数据是对其进行生态风险评价的主要依据^[2]. 实验测定生

收稿日期: 2019-09-20; 修回日期: 2019-11-11

基金项目: 国家自然科学基金资助项目(21808167, 21676203)

作者简介: 史琦玉(1994—), 女, 山西人, 硕士研究生; 通信作者: 贾青竹, 教授, jiaqingzhu@tust.edu.cn

物毒性不仅耗时费力,而且与动物伦理 3R 原则 (replacement, reduction, refinement) 相违背. 因此,亟需寻求快速高效获取生态毒性数据的方法. 根据欧洲化学品注册、评估、授权和限制条例 REACH 的规定,为避免和减少化学品对生态环境可能造成的潜在危害,鼓励使用数学模型预测工具如定量构效关系 (QSAR) 获取生态毒性数据. QSAR 方法已成为预测化学品生态毒理危害不可或缺的工具^[3-4].

许多研究者报道了苯衍生物对 *V. fischeri* 的水生毒性评估方法^[5-9]. Wang 等^[5]采用逐步回归分析法建立了 QSAR 模型,预测苯酚衍生物和苯胺衍生物(共 15 种)对 *V. fischeri* 的毒性 pEC₅₀ (EC₅₀ 取负对数),该模型统计分析结果 R^2 为 0.89. Melo 等^[6]运用连续偏最小二乘法 (PLS),预测了苯磺基羧酸酯衍生物(41 种)对 *V. fischeri* 的毒性,该 QSAR 模型的相关系数 R^2 为 0.884. Shi 等^[7]测定了苯硫酚衍生物(31 种)对 *V. fischeri* 的毒性,分别使用比较分子场分析 (CoMFA) 描述符和比较分子相似性指数分析 (CoMSIA) 描述符,采用逐步多元线性回归分析法建立模型并成功地预测了苯硫酚衍生物的毒性, R^2 (0.848 ~ 0.928) 和 Q^2 (0.516 ~ 0.652) 表明这两个模型均表现出良好的稳定性和预测能力. Liu 等^[8]基于 CoMFA 和 CoMSIA 描述符,采用偏最小二乘法分别建立两个 QSAR 模型,预测了苯磺基羧酸酯衍生物

(56 种)对 *V. fischeri* 毒性;其中,使用 CoMFA 描述符的模型 ($R^2 = 0.958$) 优于 CoMSIA 描述符的模型 ($R^2 = 0.933$). 综上,这些模型都取得了令人满意的预测结果,但是,大部分研究工作是针对单一类型的苯衍生物,同时针对多种类苯衍生物对 *V. fischeri* 的生态毒性的研究工作还很少.

近年来,本课题组提出了范数指数的概念,据此建立了一系列 QSAR 模型,并成功应用于离子液体对鼠细胞、细菌的生态毒性^[9-10]、有机物的物化性质以及对藻类、鱼类的生态毒性^[11-16]等领域.

本研究目的是建立一个 QSAR 模型预测多种苯衍生物对 *V. fischeri* 的生态毒性.

1 方法和理论

1.1 数据集

本工作中,苯衍生物生态毒性实验值 (EC₅₀, 单位用 mol/L 表示) 收集于 Wang 等^[17]、Shi 等^[7]、Aruoja 等^[18]和 Zhao 等^[19]的研究工作. 参与建模的 110 种苯衍生物包括苯酚衍生物和苯胺衍生物(共 58 种)、苯硫酚衍生物(30 种)、苯甲酸衍生物(10 种)和抗生素(12 种). 这些化合物对 *V. fischeri* 的生态毒性实验值 pEC₅₀ 等的信息列于表 1 中.

表 1 苯衍生物对 *V. fischeri* 的毒性值

Tab. 1 Toxicity of benzene derivatives to *V. fischeri*

序号	化合物英文名称	CAS	pEC ₅₀	
			实验值	计算值
1	Thiophenol	108-98-5	5.71	5.29
2	2-Fluorothiophenol	2557-78-0	4.78	4.91
3	4-Fluorothiophenol	371-42-6	4.97	4.92
4	3-Fluorothiophenol	2557-77-9	5.06	4.86
5	2,3,5,6-Tetrafluorothiophenol	769-40-4	4.86	4.59
6	Pentafluorothiophenol*	771-62-0	4.34	4.39
7	2,4-Difluorothiophenol	1996-44-7	5.15	4.83
8	2-Chlorothiophenol	6320-03-2	4.90	5.16
9	3-Chlorothiophenol*	2037-31-2	5.03	5.13
10	2,3-Dichlorothiophenol	17231-95-7	4.91	5.37
11	4-Chlorothiophenol	106-54-7	4.99	5.18
12	2,4-Dichlorothiophenol	1122-41-4	5.59	5.27
13	3,5-Dichlorothiophenol	17231-94-6	5.11	5.29
14	2,5-Dichlorothiophenol*	5858-18-4	5.16	5.25
15	2,6-Dichlorothiophenol	24966-39-0	4.99	5.15
16	3-Bromothiophenol	6320-01-0	4.57	5.05
17	4-Bromothiophenol*	106-53-6	5.60	5.09
18	2-Bromothiophenol	6320-02-1	4.88	4.87
19	2-Amino-4-chlorothiophenol	1004-00-8	5.37	5.03
20	2,4-Dimethylthiophenol	13616-82-5	4.77	4.80
21	2,5-Dimethylthiophenol	4001-61-0	4.66	4.78

续表

序号	化合物英文名称	CAS	pEC ₅₀	
			实验值	计算值
22	o-Methylthiophenol	137-06-4	4.49	4.62
23	p-Methylthiophenol	87728-89-0	5.89	4.97
24	2,6-Dimethylthiophenol	118-72-9	4.64	4.65
25	m-Methylthiophenol	108-40-7	4.60	4.78
26	3,4-Dimethylthiophenol	18800-53-8	4.95	4.98
27	2-Aminothiophenol	137-07-5	4.72	4.74
28	4-Aminothiophenol	1193-02-8	4.66	5.04
29	3-Methoxythiophenol	15570-12-4	4.26	4.40
30	3,4-Dimethoxythiophenol	700-96-9	4.72	4.87
31	Phenol	108-95-2	2.68	2.76
32	2-Chlorophenol	95-57-8	3.39	3.65
33	3-Chlorophenol	108-43-0	4.05	3.97
34	4-Chlorophenol	106-48-9	3.61	4.08
35	2,3-Dichlorophenol	576-24-9	4.17	4.21
36	2,4-Dichlorophenol	120-83-2	4.30	4.27
37	2,5-Dichlorophenol	583-78-8	4.65	4.34
38	2,6-Dichlorophenol*	87-65-0	4.01	4.04
39	3,4-Dichlorophenol	95-77-2	4.87	4.61
40	3,5-Dichlorophenol	591-35-5	4.89	4.59
41	2,3,4-Trichlorophenol	15950-66-0	4.68	4.60
42	2,3,5-Trichlorophenol	933-78-8	4.94	4.78
43	2,3,6-Trichlorophenol	933-75-5	4.39	4.38
44	2,4,5-Trichlorophenol	95-95-4	4.42	4.72
45	2,4,6-Trichlorophenol*	88-06-2	4.54	4.42
46	Aniline	62-53-3	3.24	2.88
47	2-Chloroaniline*	95-51-2	3.51	3.59
48	3-Chloroaniline	108-42-9	3.68	4.25
49	4-Chloroaniline*	106-47-8	4.56	4.41
50	2,3-Dichloroaniline	608-27-5	4.38	4.23
51	2,4-Dichloroaniline	554-00-7	4.61	4.32
52	2,5-Dichloroaniline	95-82-9	3.99	4.09
53	2,6-Dichloroaniline	608-31-1	3.84	3.97
54	3,4-Dichloroaniline	95-76-1	4.81	4.84
55	3,5-Dichloroaniline	626-43-7	4.57	4.81
56	2,3,4-Trichloroaniline*	634-67-3	4.74	4.74
57	2,4,5-Trichloroaniline	636-30-6	4.80	4.62
58	2,4,6-Trichloroaniline	634-93-5	4.60	4.43
59	3,4,5-Trichloroaniline	634-91-3	5.14	5.21
60	2-Methylaniline*	95-53-4	2.99	2.97
61	3-Methylaniline	108-44-1	3.60	3.36
62	4-Methylaniline	106-49-0	3.40	3.39
63	2,3-Dimethylaniline	87-59-2	3.59	3.48
64	2,4-Dimethylaniline	95-68-1	3.49	3.56
65	2,5-Dimethylaniline	95-78-3	3.23	3.53
66	2,6-Dimethylaniline	87-62-7	3.05	2.99
67	3,4-Dimethylaniline	95-64-7	4.22	3.74
68	3,5-Dimethylaniline	108-69-0	3.64	3.73
69	2,4,6-Trimethylaniline*	88-05-1	3.82	3.64
70	2-Ethylaniline	578-54-1	3.39	3.54
71	3-Ethylaniline	587-02-0	3.93	3.81
72	4-Ethylaniline*	589-16-2	4.14	3.85
73	2,6-Diethylaniline	579-66-8	3.56	3.88
74	2-Methylphenol	95-48-7	2.93	3.03

续表

序号	化合物英文名称	CAS	pEC ₅₀	
			实验值	计算值
75	3-Methylphenol	108-39-4	2.87	3.42
76	4-Methylphenol	106-44-5	3.27	3.42
77	2,3-Dimethylphenol*	526-75-0	3.40	3.49
78	2,4-Dimethylphenol	105-67-9	3.80	3.70
79	2,5-Dimethylphenol	95-87-4	3.58	3.47
80	2,6-Dimethylphenol*	576-26-1	3.47	3.18
81	3,4-Dimethylphenol	95-65-8	3.58	3.82
82	3,5-Dimethylphenol	108-68-9	3.65	3.92
83	2,3,5-Trimethylphenol	697-82-5	4.00	4.07
84	2,3,6-Trimethylphenol*	2416-94-6	3.98	3.75
85	2,4,6-Trimethylphenol	527-60-6	4.15	3.94
86	2-Ethylphenol	90-00-6	3.59	3.42
87	3-Ethylphenol	620-17-7	3.48	3.76
88	4-Ethylphenol	123-07-9	3.75	3.78
89	Benzoic acid	65-85-0	4.09	3.64
90	4-Chlorobenzoic acid	74-11-3	4.39	4.43
91	3-Chlorobenzoic acid	535-80-8	4.32	4.31
92	2-Chlorobenzoic acid	118-91-2	4.17	4.01
93	4-Bromobenzoic acid*	586-76-5	4.83	4.80
94	3-Bromobenzoic acid	585-76-2	4.66	4.76
95	4-Fluorobenzoic acid	456-22-4	3.96	3.97
96	4-Aminobenzoic acid	150-13-0	4.00	4.25
97	3-Aminobenzoic acid*	99-05-8	3.87	4.25
98	2-Aminobenzoic acid	118-92-3	4.00	4.25
99	Sulfamethazine	57-68-1	4.45	4.59
100	Sulfamer	651-06-9	4.36	4.74
101	Sulfamethoxypyridazine	80-35-3	4.86	4.65
102	Sulfisoxazole	127-69-5	4.63	4.60
103	Sulfadoxine	2447-57-6	4.04	4.09
104	Sulfamethoxazole*	723-46-6	4.87	4.60
105	Sulfadiazine	68-35-9	4.28	4.23
106	Sulfamonomethoxine*	1220-83-3	4.89	4.83
107	Trimethoprim*	738-70-5	6.49	6.43
108	Diaveridine	5355-16-8	6.13	5.97
109	Ormetoprim*	6981-18-6	6.51	6.41
110	Pyrimethamine*	58-14-0	5.04	5.26

注: *表示测试集; 序号 1—30 数据来源于文献[7]; 序号 31—88 数据来源于文献[18]; 序号 89—98 数据来源于文献[19]; 序号 99—110 数据来源于文献[17].

1.2 原子分布矩阵

本工作利用 HyperChem7.0 软件 (<http://www.hyper.com>) 进行结构优化. 优化方法采用量子化学从头计算方法, 在 STO-3 G 基组下优化计算, 得到化合物稳定的结构. 根据原子性质和空间分布, 建立了原子分布矩阵 (M), 并对 M 进一步数据挖掘. M 范数计算即为范数指数. 范数定义表达为式(1)一式(3).

$$\|M\|_1 = \max_j \left(\sum_i |m_{ij}| \right) \quad (1)$$

$$\|M\|_2 = \sqrt{\max(\lambda_i(M^H \times M))} \quad (2)$$

$$\|M\|_3 = \frac{1}{n} \sum_j \sum_i |m_{ij}| \quad (3)$$

式中: m_{ij} 表示矩阵中第 i 行 j 列的元素; n 表示化合物的总数; λ_i 表示矩阵的特征值; M^H 是矩阵 M 的转置共轭矩阵.

采用矩阵 (D , 式(4)) 和相邻矩阵 (A , 式(5)) 来描述原子的位置分布. 原子性质包含范德华半径 (vr)、原子电荷 (ac)、电子层数 (es)、原子质量 (aw) 和分子质量 (mw), 原子性质矩阵 (P) 表达为式(6)一式(10). 本文建立的 8 个原子分布矩阵列在表 2 中.

表2 8个原子分布矩阵
Tab. 2 8 atomic distribution matrices

序号	M	序号	M
M ₁	$P_1^T \times A$	M ₅	$P_2^T \times D \times A$
M ₂	$P_2^T \times A$	M ₆	$P_4 \times P_4^T \times D$
M ₃	$P_3^T \times D$	M ₇	$P_3 \times P_5^T \times D \times A$
M ₄	$P_5^T \times D$	M ₈	$P_4 \times P_4^T \times D \times A$

$$D = [s_{ij}] \quad s_{ij} = \begin{cases} d_{ij}^{-1} & i \neq j \\ 0 & \text{其他} \end{cases} \quad (4)$$

$$A = [s_{ij}] \quad s_{ij} = \begin{cases} 1 & \text{如果 } i \text{ 和 } j \text{ 相邻} \\ 0 & \text{其他} \end{cases} \quad (5)$$

式(4)中: d_{ij} 是原子 i 和 j 之间的欧氏空间距离. HyperChem7.0 软件中原子的欧氏空间距离以 Å 为单位, 无量纲.

$$P_1 = [aw_i] \quad (6)$$

$$P_2 = [aw_i/mw_i] \quad (7)$$

$$P_3 = [es_i] \quad (8)$$

$$P_4 = [ac_i] \quad (9)$$

$$P_5 = [ac_i/vr_i] \quad (10)$$

P_1 — P_5 为列向量. 原子的范德华半径以 Å 为单位, 无量纲.

$$\times: \text{(i). } Q = [q_j] \quad W = [w_{ij}] \quad Q \times W = [q_j \times w_{ij}];$$

$$\text{(ii). } Q = [q_{ij}] \quad W = [w_{ij}] \quad Q \times W = [q_{ij} \times w_{ij}].$$

1.3 模型验证

根据 OECD 指导原则, 良好的 QSAR 模型应具有适当的拟合度、稳健性和预测能力. QSAR 模型验证最常用的方法包括内部验证和外部验证^[20]. 其中, 统计参数为相关性系数的平方 (R^2)、Fisher 值 (F) 和均方根误差 (RMSE) 等.

在内部验证过程中, 留一法交叉验证 (LOO CV) 和五折交叉验证 (5-fold CV) 的数据拟合质量统计参数分别是 Q^2 , $Q_{5\text{-fold}}^2$ 和均方根误差 (RMSE_{LOO} 和 RMSE_{5-fold}). 为了进行外部验证, 整个数据集以 4:1 的比例分为训练集 (88 个有机物) 和测试集 (22 个有机物). 对于训练集 (tr) 和测试集 (te), 分别使用外部验证的统计参数 (R_{tr}^2 , R_{te}^2) 和均方根误差 (RMSE_{tr}, RMSE_{te}) 验证模型的预测能力. 另外, 本工作通过 Y 随机测试检查了模型的偶然相关性.

2 结果和讨论

2.1 模型建立

本文提出了一个预测多种类苯衍生物对 *V.*

fischeri 生态毒性 (pEC₅₀) 的模型, 如式 (11) 所示, 其中模型参数见表 3.

$$pEC_{50} = 13.5914 + \sum_{k=1}^8 b_k I_k \quad (11)$$

此模型: $n = 110$, $R^2 = 0.895$, $F = 108$, RMSE = 0.241, $Q^2 = 0.874$, RMSE_{LOO} = 0.264.

表3 模型中8个范数描述符及其对应系数值

Tab. 3 8 norm descriptors in the model and their corresponding coefficient values

k	b_k	I_k	k	b_k	I_k
1	-0.112 2	$\ M_1\ _2$	5	-3.829 3	$\ M_5\ _1$
2	32.226 9	$\ M_2\ _2$	6	-4.561 3	$\ M_6\ _1$
3	0.414 4	$\ M_3\ _1$	7	24.028 9	$\ M_7\ _2$
4	-32.625 4	$\ M_4\ _3$	8	-5.083 2	$\ M_8\ _1$

2.2 模型预测结果

苯衍生物对 *V. fischeri* 生态毒性的模型计算值列在表 1 中, 实验值和模型计算 pEC₅₀ 值的散点图如图 1 所示. 由图 1 可知: 实验与计算 pEC₅₀ 值吻合较好. 模型统计结果 R^2 、 F 、RMSE 分别为 0.895、108 和 0.241, 说明了该模型能够准确预测苯衍生物对 *V. fischeri* 的生态毒性.

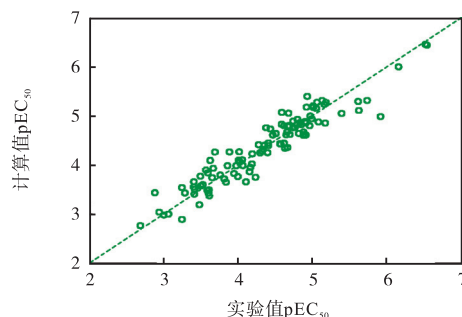


图1 实验值与计算值 pEC₅₀ 散点图

Fig. 1 Scatter diagram of experimental and calculated pEC₅₀ values

2.3 内部验证

本工作利用留一法和五折交叉验证法对模型进行内部验证. 图 2 为模型 (式 (11)) 与留一法 (LOO CV)、五折交叉验证 (5-fold CV) 计算结果的误差分布图. 图 2 结果显示, 本工作模型与留一法、五折交叉验证的误差分布具有高度一致性, 三者总体误差分布均在 $[-0.15 \sim 0.15]$ 之间, 并符合正态分布. 统计结果显示, Q_{LOO}^2 和 $Q_{5\text{-fold}}^2$ 都等于 0.874, 进一步表明本工作所建立模型具有很好的稳健性.

2.4 外部验证

训练集和测试集实验值与计算值 pEC₅₀ 的散点

图如 3 所示. 从图 3 中可以看出, 训练集和测试集的实验值和预测值都与对角线相近, 外部验证的 R_{tr}^2 、 R_{te}^2 分别为 0.873 和 0.938, $RMSE_{tr}$ 、 $RMSE_{te}$ 分别为 0.247 和 0.242, 说明基于范数描述符建立的 QSAR 模型稳定性好, 具有良好的预测能力. 此外, 其他验证参数^[20]: Q_{F1}^2 为 0.932、 Q_{F2}^2 为 0.927、 Q_{F3}^2 为 0.878、CCC 为 0.965, 均说明该模型具有良好的预测能力.

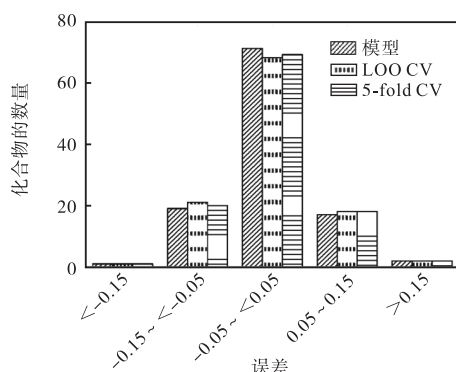


图 2 模型、留一法交叉验证和五折交叉验证的误差分布
Fig. 2 Error distribution of the new model, LOO CV and 5-fold CV

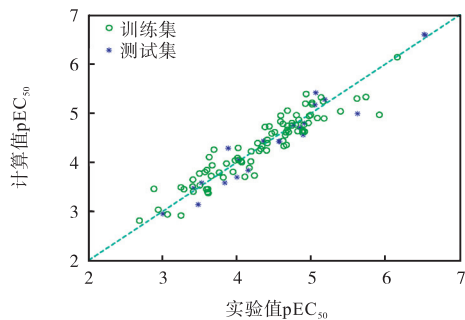


图 3 训练集和测试集实验值和计算值 pEC_{50} 散点图
Fig. 3 Scatter diagram of experimental and calculated pEC_{50} values of the training set and testing set

2.5 Y 随机验证

Y 随机化检验对 QSAR 模型的验证具有重要意义^[21]. 本工作将有机的毒性实验值 pEC_{50} 随机打乱顺序 1000 次, 对模型进行 1000 次 Y 随机化检验, 结果如图 4 所示.

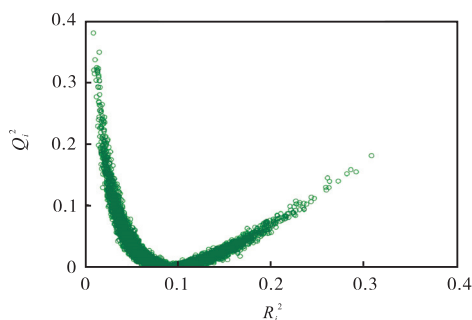


图 4 1000 次 Y 随机检验结果
Fig. 4 Results of 1000 times of Y-randomization test

图 4 结果表明, 所有 R_i^2 和 Q_i^2 均小于 0.4, 平均值分别为 0.072 和 0.028, 远远低于原模型的精度统计值 R^2 和 Q^2 . 据此推断, 本工作在建模过程中不存在偶然相关, 也不存在过度拟合的可能性, 表明模型具有较好的稳健性.

2.6 应用域分析

为了获得可靠的预测效果, 有必要验证 QSAR 模型的应用领域. 本工作利用三维 Williams plot 描述 pEC_{50} 模型的应用域. 在该图中, 杠杆 (h_i) 度量化合物离模型适用性范围的远近, 临界值 (h^*) 由 $3 \times (p + 1)/n$ 计算, 其中 p 为模型变量数, n 为建模过程中训练集的数量. 对于 h 大于 h^* 的化合物, 认为其属于结构异常的范畴; 对于交叉验证的标准残差大于 3 个标准偏差单位的化合物, 则认为是响应异常的范畴.

本工作模型应用域的验证结果如图 5 所示, 几乎所有化学物质位于可接受区域内, 仅存在一个结构异常和一个响应异常值. 由此可见, 本工作建立的模型具有广泛的应用域, 可以为苯衍生物对 *V. fischeri* 生态毒性的预测提供可靠数据.

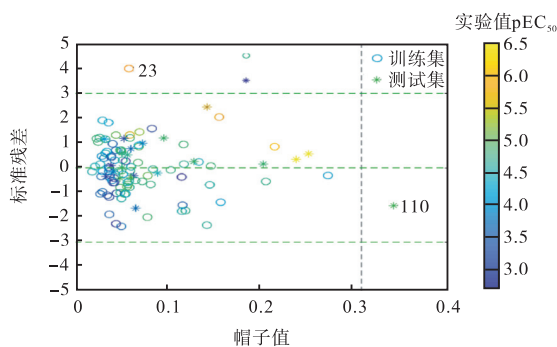


图 5 模型的应用域
Fig. 5 Applicability domain of the new model

2.7 与文献对比

将本工作模型与其他文献模型进行比较, 比较结果见表 4. 首先文献研究基本上都是针对小样本量的某一类化合物进行 QSAR 研究, 均取得较好的模型预测精度^[6, 7, 17, 22]. Guimarães 等^[22]和 Shi 等^[7]的研究工作得到了较高的 R^2 (0.848 ~ 0.928), 但是这些模型交叉验证的 Q^2 均较低, 由此说明, 这些模型的精度较高但是稳定性不高. 本工作针对多种类苯衍生物对 *V. fischeri* 的生态毒性进行建模研究, 不仅样本量大, 而且模型的稳定性高 ($Q^2 = 0.874$); 本工作基于范数描述符建立的 QSAR 模型稳定性良好, 预测能力较强, 表明该模型可以较为成功地预测苯衍生物对 *V. fischeri* 的生态毒性.

表 4 本工作与文献研究的比较
Tab. 4 Comparison of this research with references

来源	化合物类别	建模方法	统计结果	
文献[22]	苯胺衍生物和苯酚衍生物	PLS	n	58
			R^2	0.880
			RMSE	0.240
			Q^2	0.560
			RMSE _{LOO}	0.490
文献[7]	苯硫酚衍生物	CoMFA-PLS	n	31
			R^2	0.928
			F	47
			Q^2	0.652
			文献[6]	苯砒基羧酸酯衍生物
R^2	0.848			
F	21			
Q^2	0.516			
文献[17]	抗生素	MLR		
			R^2	0.884
			F	145
			RMSE	0.162
			Q^2	0.860
文献[17]	抗生素	MLR	RMSE _{LOO}	0.172
			n	18
			R^2	0.699
			F	17
			RMSE	0.161
本工作	多种苯衍生物 ^b	MLR	Q^2	0.660
			RMSE _{LOO}	0.174
			n	110
			R^2	0.895
			F	108
本工作	多种苯衍生物 ^b	MLR	RMSE	0.251
			Q^2	0.874
			RMSE _{LOO}	0.264

注: ^a 表示训练集样本的个数; ^b 表示苯衍生物包括苯酚衍生物和苯胺衍生物(共 58 种)、苯硫酚衍生物(30 种)、苯甲酸衍生物(10 种)、12 种抗生素。

3 结 论

本工作提出了范数描述符, 据此建立 QSAR 模型, 对 5 类苯衍生物(苯胺、苯酚、硫酚、苯甲酸和抗生素)的费氏弧菌生态毒性进行预测。主要研究结论如下: (1) 本工作模型对费氏弧菌生态毒性具有很好的预测精度; (2) 统计结果和模型验证结果表明了本工作模型的稳定性和广泛适用性; (3) 基于原子分布矩阵构建的范数描述符能够准确描述有机物的分子结构, 据此建立的模型对于有机物的生态风险评价具有重要意义。

参考文献:

- [1] Huang H, Wang X D, Ou W H, et al. Acute toxicity of benzene derivatives to the tadpoles (*Rana japonica*) and QSAR analyses[J]. Chemosphere, 2003, 53 (8): 963–970.
- [2] Vlastos D, Antonopoulou M, Konstantinou I. Evaluation of toxicity and genotoxicity of 2-chlorophenol on bacteria, fish and human cells[J]. Science of The Total Environment, 2016, 551: 649–655.
- [3] Toropov A A, Raška Jr I, Toropova A P, et al. The study of the index of ideality of correlation as a new criterion of predictive potential of QSPR/QSAR-models[J]. Science of the Total Environment, 2019, 659: 1387–1394.
- [4] Ghanem O B, Mutalib M I A, Lévêque J M, et al. Development of QSAR model to predict the ecotoxicity of *Vibrio fischeri* using COSMO-RS descriptors[J]. Chemosphere, 2017, 170: 242–250.
- [5] Wang X H, Fan L Y, Wang S, et al. Relationship between acute and chronic toxicity for prevalent organic pollutants

- in *Vibrio fischeri* based upon chemical mode of action[J]. *Journal of Hazardous Materials*, 2017, 338: 458–465.
- [6] de Melo E B, Martins J P A, Miranda E H, et al. A best comprehension about the toxicity of phenylsulfonyl carboxylates in *Vibrio fischeri* using quantitative structure activity/property relationship methods[J]. *Journal of Hazardous Materials*, 2016, 304: 233–241.
- [7] Shi J Q, Cheng J, Wang F Y, et al. Acute toxicity and *n*-octanol/water partition coefficients of substituted thiophenols: Determination and QSAR analysis[J]. *Ecotoxicology and Environmental Safety*, 2012, 78: 134–141.
- [8] Liu X H, Yang Z F, Wang L S. Three-dimensional quantitative structure-activity relationship study for phenylsulfonyl carboxylates using CoMFA and CoMSIA[J]. *Chemosphere*, 2003, 53 (8) : 945–952.
- [9] Yan F Y, Lan T, Yan X, et al. Norm index-based QSTR model to predict the eco-toxicity of ionic liquids towards Leukemia rat cell line[J]. *Chemosphere*, 2019, 234 : 116–122.
- [10] Yan F Y, He W S, Jia Q Z, et al. QSAR models for describing the toxicological effects of ILs against *Candida albicans* based on norm indexes[J]. *Chemosphere*, 2018, 201 : 417–424.
- [11] Yan F Y, Liu T T, Jia Q Z, et al. Multiple toxicity endpoint-structure relationships for substituted phenols and anilines[J]. *Science of The Total Environment*, 2019, 663 : 560–567.
- [12] Wang Y L, Yan F Y, Jia Q Z, et al. Distributive structure-properties relationship for flash point of multiple components mixture[J]. *Fluid Phase Equilibria*, 2018, 474: 1–5.
- [13] Jia Q Z, Yan X, Lan T, et al. Norm indexes for predicting enthalpy of vaporization of organic compounds at the boiling point[J]. *Journal of Molecular Liquids*, 2019, 282: 484–488.
- [14] Jia Q Z, Zhao Y P, Yan F Y, et al. QSAR model for predicting the toxicity of organic compounds to fathead minnow[J]. *Environmental Science and Pollution Research*, 2018, 25 (35) : 35420–35428.
- [15] 崔雪, 贾青竹, 李磊, 等. 有机物在水中溶解度范数指数法定量构效关系[J]. *天津科技大学学报*, 2016, 31 (4) : 35–39.
- [16] 刘英, 尹京晨, 计惠芬, 等. 基于范数指数构效关系预测单端孢霉烯真菌毒素毒性[J]. *天津科技大学学报*, 2018, 33 (2) : 36–41.
- [17] Wang D L, Wu X D, Lin Z F, et al. A comparative study on the binary and ternary mixture toxicity of antibiotics towards three bacteria based on QSAR investigation[J]. *Environmental Research*, 2018, 162: 127–134.
- [18] Aruoja V, Sihtmäe M, Dubourguier H C, et al. Toxicity of 58 substituted anilines and phenols to algae *Pseudokirchneriella subcapitata* and bacteria *Vibrio fischeri*: Comparison with published data and QSARs[J]. *Chemosphere*, 2011, 84 (10) : 1310–1320.
- [19] Zhao Y H, Ji G D, Cronin M T D, et al. QSAR study of the toxicity of benzoic acids to *Vibrio fischeri*, *Daphnia magna* and carp[J]. *Science of the Total Environment*, 1998, 216 (3) : 205–215.
- [20] Roy K, Kar S, Das R N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment[M]. Boston: Academic Press, 2015: 231–289.
- [21] Rücker C, Rücker G, Meringer M. γ -Randomization and its variants in QSPR/QSAR[J]. *Journal of Chemical Information and Modeling*, 2007, 47 (6) : 2345–2357.
- [22] Guimarães M C, da Mota E G, Silva D G, et al. aug-MIA-QSPR modelling of the toxicities of anilines and phenols to *Vibrio fischeri* and *Pseudokirchneriella subcapitata*[J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, 134: 53–57.

责任编辑: 周建军