

DOI:10.13364/j.issn.1672-6510.20160377

## 基于范数指数构效关系预测单端孢霉烯真菌毒素毒性

刘 英<sup>1</sup>, 尹京晨<sup>1</sup>, 计惠芬<sup>2</sup>, 贾青竹<sup>1</sup>

(1. 天津科技大学海洋与环境学院, 天津 300457; 2. 浙江东洋环境工程有限公司, 湖州 313000)

**摘要:** 依据本课题组前期提出的范数指数描述符, 建立了一个预测单端孢霉烯真菌毒素(trichothecene mycotoxins)毒性的 QSAR 模型, 并对 35 个毒素物质的毒性进行了计算. 留一交叉验证和 Y 随机验证表明这个新计算模型预测结果可靠、稳定( $R^2$  为 0.944 2,  $Q^2$  为 0.878 9); 同时, 该模型的应用域验证结果表明此模型有可能在大范围上推广运用. 模型统计结果和相关验证结果都表明基于范数指数建立的 QSAR 模型可以成功应用于预测单端孢霉烯真菌毒素毒性.

**关键词:** 单端孢霉烯真菌毒素; 毒性; 范数指数; 定量构效关系; 留一交叉验证; 应用域

中图分类号: Q936; O621.1 文献标志码: A 文章编号: 1672-6510(2018)02-0032-06

## Structure-activity Relationship for Predicting the Toxicity of Trichothecene Mycotoxins Based on Norm Indexes

LIU Ying<sup>1</sup>, YIN Jingchen<sup>1</sup>, JI Huifen<sup>2</sup>, JIA Qingzhu<sup>1</sup>

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

2. Zhejiang Dongyang Environmental Engineering Co., Ltd., Huzhou 313000, China)

**Abstract:** Based on the norm indexes descriptor proposed earlier in this project, a model for predicting the toxicity of 35 trichothecene mycotoxins was developed. The model was validated by leave-one-out validation and Y-randomization test with satisfactory results ( $R^2$  of 0.944 2,  $Q^2$  of 0.878 9), which further demonstrated that this model is reliable and stable. Besides, the applicability domain of this model was validated by using the leverage approach and the results suggested a potentially large scale of further utilization. The statistical values and validation test results demonstrated that our proposed norm indexes-based model could be successfully applied to predicting the toxicity of trichothecene mycotoxins.

**Key words:** trichothecene mycotoxins; toxicity; norm indexes; quantitative structure-activity relationship; leave-one-out validation; applicability domain

单端孢霉烯族毒素(trichothecenes)是一类化学性质相近的真菌毒素,由镰刀菌属、单端孢属和头孢霉属等真菌产生<sup>[1-2]</sup>,其结构属于倍半萜环氧化物,具有 1 个三环核心,在 C12 和 C13 位含有 1 个环氧化物,这是一个毒性必需基团(核心结构见图 1). 该类毒素不仅污染小麦和玉米等禾谷类作物,也会通过食物链危害肉、奶、蛋等畜产品,WHO 已将其列为天然存在的最危险的食物污染物之一<sup>[3]</sup>. 而且,由于该类毒素对真核细胞具有多重抑制作用,通过对蛋白质、DNA 和 RNA 合成的抑制,对线粒体功能、细胞分

裂和膜功能的抑制以及促进凋亡基因的表达,对人畜健康产生免疫抑制,导致其具有典型的“致癌、致畸、致突变”作用,从而严重威胁人畜健康<sup>[4]</sup>.

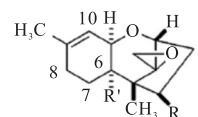


图 1 单端孢霉烯核结构

Fig. 1 Structure of the trichothecene nucleus

定量构效关系(quantitative structure-activity rela-

收稿日期: 2016-11-18; 修回日期: 2017-01-12

基金项目: 国家自然科学基金资助项目(21306137, 21676203)

作者简介: 刘 英(1990—),女,山西人,硕士研究生;通信作者: 贾青竹,教授, jiaqingzhu88@126.com

tionship, QSAR)是研究化合物结构与生物活性相关性的典型方法,广泛应用于新药物设计以及对于毒性机制的理解探究领域<sup>[5-8]</sup>。尽管已有学者对单端孢霉烯族毒素的结构与活性关系进行了考察,但是至今关于这一系列毒素的QSAR研究依然很少<sup>[5-6]</sup>。Steinmetz等<sup>[6]</sup>基于分子比较力场分析(comparative molecular field analysis, CoMFA)方法和比较分子相似因子分析(comparative molecular similarity indices analysis, CoMSIA)方法对32个大环内酯物单端孢霉烯族毒素和15个单端孢霉烯族毒素的毒性进行了3D(3 dimension)QSAR研究,结果表明单端孢霉烯族毒素类物质的取代基以及分子构象对其生物活性都有一定影响。

本课题组之前提出系列范数指数描述符已经成功地预测了有机物多种活性,包括麻醉性污染物的水生毒性、杂环化合物的药理学和毒理学活性等<sup>[7-10]</sup>。先前的研究工作表明该系列范数描述符也许可以从本质上体现分子结构,从而反映化合物的多个物化性质。

在本课题组之前工作的基础上,本工作又提出新的描述符,并据此建立单端孢霉烯族真菌毒素的毒性预测模型,对35个单端孢霉烯族真菌毒素的毒性进行了计算,采用留一法交叉验证(LOO)、外部验证和Y随机验证手段对模型进行了验证,并对模型的应用域(AD)进行了评价。本工作深入研究此类化合物分子结构对其毒性的影响,将有助于人们采取更科学的毒素防控及清除机制。

## 1 材料与方法

### 1.1 样本集

Jarvis等<sup>[11-12]</sup>测定了大环内酯物真菌毒素在延长接种p-388白血病小鼠生命中的能力,这些数据可以从根本上反映大环内酯物真菌毒素对于小鼠白血病细胞的影响,并将测定活性定量为R,即 $R = 100 \times$ 测试集动物存活天数/对照组动物存活天数。在此工作中,包含大环内酯物真菌毒素毒性数据的化合物分子(代表性分子结构见图2)从文献中<sup>[6,13-14]</sup>获得。

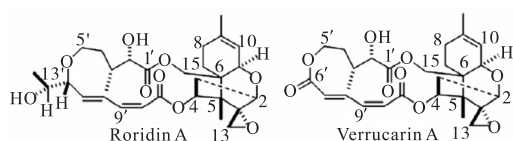


图2 两种大环内酯物真菌毒素结构

Fig. 2 Structure of two macrolide trichothecene mycotoxins

### 1.2 分子结构优化

利用软件HyperChem 7.0构建分子的3D结构,对有机物分子结构进行优化,具体采用量子化学中的从头算法 $ab\ initio$ 在ST0-3G进行能量最低优化,分子结构信息中包括分子中原子的种类、数量、分支度,原子之间的键连接形式以及稳定结构下的原子电荷。

### 1.3 范数指数描述符和模型

在分子结构优化基础上,为了区分原子的特点和种类,本工作提出了性质矩阵。同时,欧氏空间距离矩阵是一种能够全面、有效的描述分子的方法,因此,本工作提出了不同的矩阵。

下面列出上述具体矩阵。

$$T_1 = (s_{ij})$$

$$s_{ij} = \begin{cases} 1/n & \text{原子}i\text{和}j\text{之间的跨度为}n \\ 0 & \text{其他} \end{cases} \quad (1)$$

$$T_2 = (a_{ij}) \quad a_{ij} = \begin{cases} 1 & \text{原子}i\text{和}j\text{相邻} \\ 0 & \text{其他} \end{cases} \quad (2)$$

性质矩阵:

$$P_E = \begin{bmatrix} 1/\text{电负性} \\ \text{原子质量}/\text{分子质量} \\ \text{原子电荷} \\ \log_{10}(\text{支化度}+1) \end{bmatrix} \quad (3)$$

为了讨论上述矩阵的模,本工作使用了两类范数指数 $\text{norm}(TP, 1)$ 和 $\text{norm}(TP, 2)$ ,其定义式如下:  
 $\text{norm}(TP, 1)$ : Fro 函数

$$\|TP\|_1 = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |TP_{ij}|^2} \quad (4)$$

$\text{norm}(TP, 2)$ : 矩阵的最大特征值

$$\|TP\|_2 = \left( \max \{ \lambda_j(TP \times TP^H) \} \right)^{1/2} \quad (5)$$

将矩阵 $T_b$ ( $b = 1, 2$ )和性质矩阵 $P_E$ ( $a = 1, 2, 3, 4$ )进行不同形式的组合,形成新矩阵,其定义如下:

$$TP_{1,a,b} = [P_E(:, a) \quad T_b] \quad (6)$$

$$TP_{2,a,b} = [P_E(:, a) \times P_E(:, a)^T + T_b] \quad (7)$$

$$TP_{3,a,b} = [(P_E(:, a) \times P_E(:, a)^T) \times T_b] \quad (8)$$

在式(8)中,操作符 $(\cdot \times)$ 的含义是标量相乘。通过范数计算得到48个范数描述符,为避免描述符过多而造成模型过度拟合,对建模所用描述符进行优化筛选,最后优选了6个范数描述符,建立预测模型如下:

$$\log_{10} R = 21.1589 + \sum_{i=1}^6 b_i \text{norm}(TP) \quad (9)$$

$$R^2 = 0.9442, Q^2 = 0.8789, R^2_{\text{training}} = 0.9254,$$

$$F = 78.95, R^2_{\text{test}} = 0.9966$$

模型中系数  $b$  见表 1.

表 1 模型系数

Tab. 1 Parameters of this model

序号	Norm	$b_i$
$i = 1$	$\text{norm}(TP_{1,3,2}, 1)$	20.78
$i = 2$	$\text{norm}(TP_{2,1,2}, 1)$	6.90
$i = 3$	$\text{norm}(TP_{2,4,2}, 1)$	-19.54
$i = 4$	$\text{norm}(TP_{3,1,1}, 1)$	-38.85
$i = 5$	$\text{norm}(TP_{3,3,1}, 1)$	377.71
$i = 6$	$\text{norm}(TP_{2,4,1}, 2)$	2.66

### 1.4 模型评价

通过下列参数评价本工作建立模型的稳定性: 相关系数的平方 ( $R^2$ )、平均相对误差  $\bar{\delta}$ 、 $F$  值. 为了进一步验证模型的准确性, 还使用了留一交叉验证法 ( $Q^2$ )、外部验证法 ( $R^2_{\text{training}}, R^2_{\text{test}}$ )、 $Y$  随机验证法进行验证.

$$R^2_{\text{training}} = 1 - \frac{\sum (Y_{\text{obs}(\text{train})} - Y_{\text{pre}(\text{train})})^2}{\sum (Y_{\text{obs}(\text{train})} - \bar{Y}_{\text{train}})^2} \quad (10)$$

$$R^2_{\text{test}} = 1 - \frac{\sum (Y_{\text{obs}(\text{test})} - Y_{\text{pre}(\text{test})})^2}{\sum (Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{test}})^2} \quad (11)$$

$$F = \frac{[\sum (Y_{\text{pre}} - \bar{Y})^2 - \sum (Y_{\text{pre}} - Y_{\text{obs}})^2] / k}{\sum (Y_{\text{pre}} - Y_{\text{obs}})^2 / (n - k - 1)} \quad (12)$$

$$Q^2 = 1 - \frac{\sum (Y_{\text{obs}(\text{train})} - Y_{\text{LOO-pre}(\text{train})})^2}{\sum (Y_{\text{obs}(\text{train})} - \bar{Y}_{\text{train}})^2} \quad (13)$$

$$\bar{\delta}(\%) = \frac{1}{N} \sum_n \frac{Y_{\text{obs}} - Y_{\text{pre}}}{Y_{\text{obs}}} \times 100 \quad (14)$$

式中:  $Y_{\text{obs}(\text{train})}$  为训练集实验值;  $Y_{\text{pre}(\text{train})}$  为训练集预测值;  $\bar{Y}_{\text{train}}$  为训练集实验值的平均值;  $Y_{\text{obs}(\text{test})}$  为测试集实验值;  $Y_{\text{pre}(\text{test})}$  为测试集预测值;  $\bar{Y}_{\text{test}}$  为测试集实验值的平均值;  $Y_{\text{obs}}$  为样本集实验值;  $Y_{\text{pre}}$  为样本集预测值;  $n$  为样本集数量;  $k$  为变量数量;  $\bar{Y}$  为样本集实验值的平均值;  $Y_{\text{LOO-pre}(\text{train})}$  为样本集留一交叉验证的预测值.

同时, 本工作选择了应用域验证本模型的应用范围<sup>[15]</sup>, 其定义为式 (15):

$$h = x \times (x^T \times x^{-1}) \times x^T \quad (15)$$

在上述等式中,  $h$  是杠杆值,  $x$  是变量值.

通过与阈值 ( $h^*$ ) 比较, 如果  $h < h^*$ , 表明在这个区

域内模型的预测性是稳定的; 反之, 这个模型的预测能力是较差的<sup>[16-17]</sup>. 阈值的定义为式 (16)

$$h^* = 3p' / n \quad (16)$$

式中:  $p'$  是自变量数量加 1;  $n$  是训练集数量.

## 2 结果与讨论

### 2.1 大环内酯物真菌毒素毒性的 QSAR 模型

利用新建模型式 (9) 对 35 个大环内酯物真菌毒素的毒性进行了预测, 图 3 是大环内酯物真菌毒素毒性的实验值  $\log_{10}R$  与预测值对比散点图.

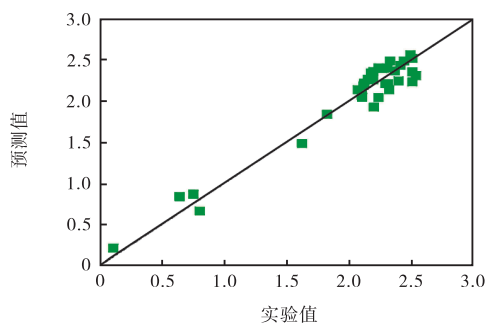


图 3  $\log_{10}R$  预测值和实验值相关性  
Fig. 3 Correlation between  $\log_{10}R$  predicted by this model and the experimental data

由图 3 可知: 所有毒素的毒性预测点和实验点均位于对角线上及附近, 表明本模型计算结果与实验值有很好的 consistency. 本预测模型相关统计数据  $R^2_{\text{training}}$ 、 $R^2_{\text{test}}$  和  $F$  值分别为 0.9254、0.9966 和 78.95, 说明本模型计算结果的精确性较好.

图 4 是残差与实验值的对比图. 从图 4 可以看出所有毒素的毒性预测残差都分布在  $-0.3 \sim 0.3$ , 这表明本工作的系统误差小, 模型预测能力稳定; 同时, 可以说明本工作提出的变量能有效地反映分子结构与毒性之间的关系.

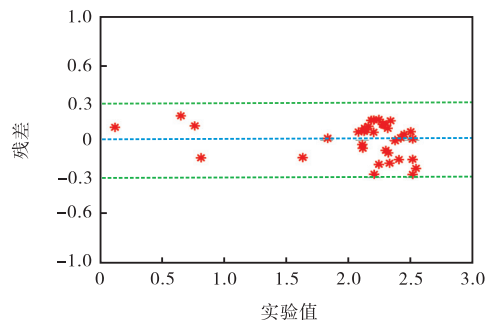


图 4 残差与实验值对比图  
Fig. 4 Plot of the residual vs. experimental data from this model

本工作与 Steinmetz 等<sup>[6]</sup>对比结果见表 2. 由表 2 可以看出文献中 CoMSIA 方法和本工作得到的相关系数  $R^2$  都大于 0.900,  $R^2$  均在同一数量级,但是本工作交叉验证的相关系数远大于 CoMSIA 方法,说明本工作的模型更稳定. 对于  $F$  值而言,本工作的  $F$  值同样远大于文献中两种方法得出的  $F$  值,更进一步证明本模型的准确性高、稳定性好.

表 2 本方法和文献方法预测结果对比

Tab. 2 Statistics comparison of reference methods and this work

	模型方法		
	本工作	The CoMFA	The CoMSIA
样本数	29/6	26/5	26/5
$R^2_{\text{training}}$	0.925	0.827	0.859
$R^2_{\text{test}}$	0.996	0.653	0.911
$R^2$	0.944	0.868	0.916
$Q^2$	0.878	0.498	0.633
$F$	78.95	15.90	13.31

本工作采用多元线性回归分析法,建立关于大环内酯物真菌毒素的结构与毒性关系模型. Steinmetz 等<sup>[6]</sup>利用 CoMFA 方法研究结果表明,空间位阻和静电因素对毒素的活性有重大影响,其中静电因素的贡献更大;同时,相比之下包含氢键供体/受体的 CoMSIA 方法计算结果更为精确. 本工作在计算描述符的优化确定中,首先利用欧氏距离以及相关距离矩阵考虑了大环内酯物真菌毒素分子中各个组成原子的空间分布;同时,为了实现大环内酯物真菌毒素分子中不同原子对其毒性的贡献影响,提出了原子属性矩阵即增广矩阵,其中包括电负性、原子质量、原子支化度和原子电荷. 由此,根据以上相关矩阵建立的范数描述符能从分子水平上反映真菌毒素分子结构特点以及真菌毒素与生物体之间的相互作用力. 综上,与前人研究相比,本工作优点在于:能给出确定的模型形式,方便其他研究者验证或使用;模型预测准确性和稳定性强.

在此以第 13 个化合物为例,计算该化合物的毒性. 化合物 Baccharin B8 的结构如图 5 所示.

首先,按照式(1—3)的定义式,根据化学图计算出该化合物的矩阵  $T_1$ 、 $T_2$  和性质矩阵  $P_E$ ; 然后依据式(6—8),计算出增广矩阵  $TP_{1,a,b}$ 、 $TP_{2,a,b}$  和  $TP_{3,a,b}$ . 表 3 是上述所有矩阵的范数描述符的值. 应用式(9)以及表 2 和表 3 中的参数,第 13 个化合物的毒性预测值计算过程如下:

$$\log_{10} R = 27.341 + 20.78 \times 9.593 + 2.66 \times 10.312 + 6.90 \times 20.251 -$$

$$19.54 \times 9.545 - 38.85 \times 5.280 + 377.71 \times 0.008 = 2.363$$

第 13 个化合物的  $\log_{10} R$  预测值是 2.363, 实验值是 2.367, 相对偏差是 0.17%.

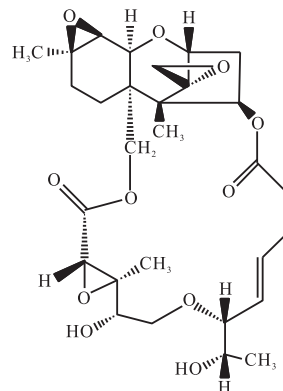


图 5 Baccharin B8 的结构

Fig. 5 Structure of Baccharin B8

表 3 第 13 个化合物的相关变量值

Tab. 3 Variable values of the thirteenth chemical

序号	norm	变量值
$i = 1$	norm( $TP_{1,3,2}, 3$ )	9.593
$i = 2$	norm( $TP_{2,1,2}, 3$ )	20.251
$i = 3$	norm( $TP_{2,4,2}, 3$ )	9.545
$i = 4$	norm( $TP_{3,1,1}, 3$ )	5.280
$i = 5$	norm( $TP_{3,3,1}, 3$ )	0.008
$i = 6$	norm( $TP_{2,4,1}, 2$ )	10.312

## 2.2 留一交叉验证

留一交叉验证法就是从所有的数据中每次选择一个分子作为一次测试集,用剩余的分子建立模型预测测试集分子的毒性. 如果留一交叉验证的相关系数  $Q^2_{\text{loo}}$  大于 0.5, 则认为模型是合理的<sup>[18]</sup>. 图 6 表明: 留一交叉验证法的毒性预测值与实验值有较好吻合度, 表明本工作建立的模型的内部预测能力较高.

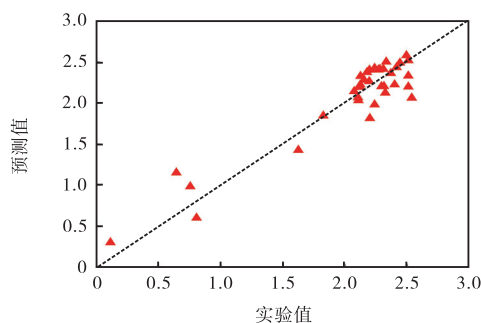


图 6 留一交叉验证预测值和实验值相关性

Fig. 6 Correlation between the data predicted by LOO-CV and the experimental data

图 7 是两种模型预测结果的标准误差 (SE) 分布



情况对比图. 由图 7 可知: 交叉验证的误差趋势和本方法建立的模型的误差趋势一致, 说明本工作建立的模型稳定、预测能力好. 同时, 留一交叉验证法的结果具有较高  $Q^2$  值(0.878 9), 以上所有验证结果均表明本工作提出的方法可以稳定地预测大环内酯物真菌毒素的毒性.

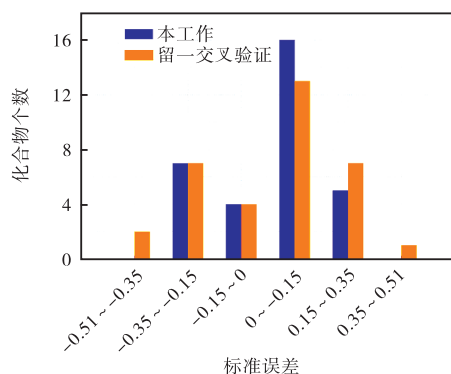


图 7 本模型和 LOO 模型毒性预测标准误差分布

Fig. 7 Standard error distributions of the toxicity predicted by this model and LOO-CV

### 2.3 外部验证

模型的内部验证只能说明该方法对于类似物的活性预测能力好, 如果用该方法预测新的不同种类的化合物的活性, 则结果有可能是不可靠的. 因此, 需要进行外部验证. 外部验证就是将全部的分子分为训练集和测试集, 训练集用于建立模型, 测试集用来验证模型. 有效的外部验证能确保本工作提出的方法对测试集分子预测的可预见性和适用性.

在本工作中, 训练集和测试集的分子与文献[6]相同, 训练集和测试集的分子毒性预测值见图 8, 其结果  $R^2_{\text{training}}$  是 0.925 4、 $R^2_{\text{test}}$  是 0.996 6. 这表明预测值和实验值有很好的一致性, 进而说明本工作建立的模型不仅可以很好地预测同类化合物毒性, 而且对多种类化合物具有较高的预测性.

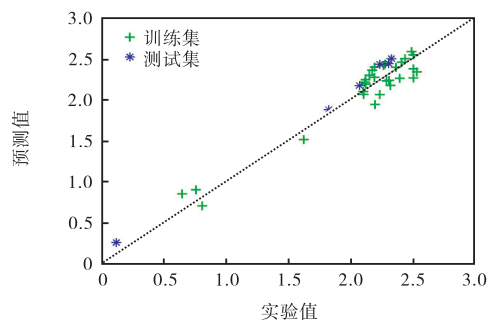


图 8 训练集和测试集的实验值和预测值的散点图

Fig. 8 Scatter plot predicted and experimental data of the training set and test set

### 2.4 Y 随机验证

Y 随机验证的原理是将实验值随机扰乱、变量不变, 再建立模型, 得到的结果与最初建立模型的结果进行对比.

在本工作中, 进行了 5 次随机验证, 新模型预测  $R^2$  和  $Q^2$  的结果与原模型的对比结果见表 4. 由表 4 可以得出, 5 次 Y 随机验证得出的新模型的  $R^2$  和  $Q^2$  都很低, 甚至为 0, 说明新模型的预测能力低. 由此证明, 本工作建立的原始模型是稳定的, 并非偶然建立.

表 4 Y 随机验证结果

Tab. 4 Results of Y randomization test of this model

验证次数	$R^2$	$Q^2$
1	0.001	0.000
2	0.002	0.001
3	0.001	0.000
4	0.000	0.000
5	0.001	0.000

### 2.5 应用域验证

对于构效关系模型统计参数的验证来说, 模型的应用域评价依然是最困难的一个评价标准. 应用域评价是模型能否准确预测新化合物性质或活性的重要指标. 目前, 已经有多种关于模型应用域评价的方法<sup>[19-21]</sup>. 本工作利用杠杆方法检测计算模型的应用域, 结果如图 9 所示. 从图 9 可以看出, 对于 35 个有机物, 只有 1 个有机物的杠杆值小于 -3, 其余所有分子的杠杆值均在规定的阈值内, 说明本方法建立的模型应用域较广泛.

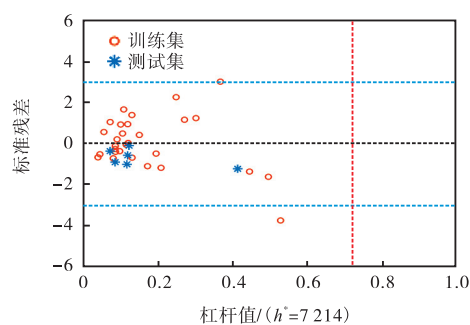


图 9 本模型的应用域

Fig. 9 Application domain of this model

## 3 结 语

本工作基于有机毒素化学结构构建的距离矩阵和性质矩阵, 提出了一个新范数描述符, 并据此建立了预测单端孢霉烯真菌毒素毒性的 QSAR 模型. 研

究结果表明:本工作提出的范数描述符能够有效描述化合物毒性与其结构之间的关系;本模型毒性预测值与实验值有很好的一致性,其中模型的相关性系数 $R^2$ 均大于0.9, $F$ 值为78.95,留一交叉验证测试( $Q^2$ 为0.8789)、 $Y$ 随机验证以及与文献的对比结果均证明本模型稳定性更好、预测能力更高、准确性更好。同时,应用域验证表明本模型计算结果准确、稳定、可靠、适用范围广。因此,本工作提出的方法将可能推广用于化学品的环境风险评价、生态健康评价等领域。

#### 参考文献:

- [1] Zhou T, He J W, Gong J. Microbial transformation of trichothecene mycotoxins[J]. World Mycotoxin Journal, 2008, 1(1): 23-30.
- [2] 邹忠义, 贺稚非, 李洪军, 等. 单端孢霉烯族毒素转化降解研究进展[J]. 食品科学, 2010, 31(19): 443-448.
- [3] 周闯, 何成华, 司慧民, 等. 2012年国内饲料及原料霉菌毒素污染调查分析[J]. 畜牧与兽医, 2014, 46(1): 81-84.
- [4] 许伟, 耿芳芳, 范梦雪, 等. 脱氧雪腐镰刀菌烯醇毒性的研究进展[J]. 生物学杂志, 2016, 33(1): 78-81.
- [5] Betina V. Structure-activity relationships among mycotoxins[J]. Chemo-Biological Interactions, 1989, 71(2/3): 105-146.
- [6] Steinmetz W E, Rodarte C B, Lin A. 3D QSAR study of the toxicity of trichothecene mycotoxins[J]. European Journal of Medicinal Chemistry, 2009, 44(11): 4485-4489.
- [7] Zhu Z, Wang Q, Jia Q, et al. Structure-property relationship for the pharmacological and toxicological activity of heterocyclic compounds[J]. Acta Physico-chimica Sinica, 2014, 30(6): 1086-1090.
- [8] Qian H, Kanwal S, Jia Q, et al. Norm index-based quantitative structure-activity relationship to predict  $\beta$ -cyclodextrin complex binding constants[J]. Acta Physico-chimica Sinica, 2015, 31(5): 893-898.
- [9] Wang Q, Jia Q, Yan L, et al. Quantitative structure-toxicity relationship of the aquatic toxicity for various narcotic pollutants using the norm indexes[J]. Chemosphere, 2014, 108: 383-387.
- [10] Jia Q, Cui X, Li L, et al. A Quantitative structure-activity relationship for high affinity 5-HT1A receptor ligands based on norm indexes[J]. Journal of Physical Chemistry B, 2015, 119(51): 15561-15567.
- [11] Jarvis B B, Stahly G P, Pavanadasivam G. Isolation and characterization of the trichoverroids and new roridins and verrucarins[J]. Journal of Medicinal Chemistry, 1982, 47(6): 1117-1124.
- [12] Jarvis B B, Vrudhula V M, Midiwo J O, et al. New trichoverroids from myrothecium verrucaria: Verrol and 12, 13-deoxytrichodermediene[J]. Journal of Medicinal Chemistry, 1983, 48(15): 2576-2578.
- [13] Cramer R D, Patterson D E, Bunce J D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins[J]. Journal of the American Chemical Society, 1988, 110(18): 5959-5967.
- [14] Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity [J]. Journal of Medicinal Chemistry, 1994, 37(24): 4130-4146.
- [15] Netzeva T I, Worth A P, Aldenberg T, et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships[J]. American Theological Library Association, 2005, 33(2): 155-173.
- [16] Gramatica P. Principles of QSAR models validation: Internal and external[J]. QSAR & Combinatorial Science, 2007, 26(5): 694-701.
- [17] Gramatica P, Giani E, Papa E. Statistical external validation and consensus modeling: A QSPR case study for Kocprediction[J]. Journal of Molecular Graphics and Modelling, 2007, 25(6): 755-766.
- [18] Mihalić Z, Trinajstić N. A graph-theoretical approach to structure-property relationships[J]. Journal of Chemical Education, 1992, 69(9): 701.
- [19] Tropsha A. Best practices for QSAR model development, validation, and exploitation[J]. Molecular Informatics, 2010, 29(6/7): 476-488.
- [20] Tropsha A, Golbraikh A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening[J]. Current Pharmaceutical Design, 2007, 13(34): 3494-3504.
- [21] Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicability domain estimation by projection of the training set descriptor space: A review[J]. American Theological Library Association, 2005, 33(5): 445-459.

责任编辑: 郎婧