

分子全息 QSAR 方法预测苯衍生物对蝌蚪的急性毒性

黄宏¹, 杨红¹, 樊伟², 王连生³

(1. 上海水产大学海洋学院, 上海 200090; 2. 中国水产科学研究院东海水产研究所, 上海 200090; 3. 南京大学环境学院, 南京 210093)

摘要: 采用半静态培养法测定了一系列苯衍生物对蝌蚪的 12 h 致死毒性. 应用分子全息 QSAR 方法构建了化合物的亚结构与毒性之间的定量模型, 探讨了分子全息长度、碎片尺寸以及区分常数对模型质量的影响, 并检验了模型的稳健性. 结果显示, 研究的 51 种苯衍生物的半致死毒性 ($12\text{h-log}I/LC_{50}$, mol/L) 分布在 2.07 ~ 4.56 之间. 最优 HQSAR 模型以 6 ~ 7 为碎片尺寸, 83 为全息长度, 6 个主成分, Leave-One-Out 交叉验证的 r_{cv}^2 是 0.849, 最终的非交叉验证的拟合模型 r^2 为 0.942, 表明模型稳健性能良好, 可用于预测苯衍生物的毒性.

关键词: 苯衍生物; 蝌蚪; 急性毒性; 全息定量结构-活性相关性

中图分类号: X174 文献标识码: A 文章编号: 0250-3301(2005)03-0025-04

Hologram Quantitative Structure Activity Relationship for Predicting Acute Toxicity of Benzene Derivatives to the Tadpoles (*Rana japonica*)

HUANG Hong¹, YANG Hong¹, FAN Wei², WANG Lian-sheng³

(1. College of Marine Science and Technology, Shanghai Fisheries University; Shanghai 200090, China; 2. East China Sea Fisheries Science Research Institute, Chinese Academy of Fisheries Science, Shanghai 200090, China; 3. School of the Environment, Nanjing University, Nanjing 210093, China)

Abstract: Acute lethal toxicity, the negative logarithm of 12 h acute median lethal molar concentration (expressed as $12\text{h-log}I/LC_{50}$, mol/L) of a series of benzene derivatives to *Rana japonica* tadpoles was determined. The relationship between the structure of benzene derivatives and their acute lethal toxicity was investigated by using hologram quantitative structure-activity relationship (HQSAR). The influence of hologram length, fragment size and distinction parameters on the quality of HQSAR model was studied. The robustness and predictive ability of the model were also validated by Leave-One-Out cross-validation procedure. The tested 51 compounds revealed a range of acute median lethal toxicity ($12\text{h-log}I/LC_{50}$, mol/L) from 2.07 to 4.56. As a result, the best model was generated using a fragment size of 6 ~ 7 from a hologram length of 83 with 6 components. The HQSAR model showed a conventional correlation coefficient r^2 of 0.942 and a Leave-One-Out cross-validation r_{cv}^2 equals to 0.849, indicating the model possess high statistical quality in the prediction of acute toxicity of novel benzene analogs.

Key words: benzene derivatives; tadpoles (*Rana japonica*); acute toxicity; HQSAR

苯及其衍生物对环境的潜在威胁很大^[1]. 很多实验室测定过处于食物链的不同等级上的水生生物的急性毒性, 但主要集中在鱼类或贝类生物^[2, 3]. 青蛙是典型的水陆两栖动物, 是水生生物与陆生动物连接的纽带, 并且青蛙在农业上也有重要地位, 研究有毒污染物对青蛙的幼体——蝌蚪的毒性效应, 进而保护青蛙十分必要.

在过去的几十年中, QSAR 研究已经被广泛用于预测未测试的化合物的毒性^[4]. 而大多数成功的 QSAR 预测模型都是建立在化合物相同或相似的毒性作用机制上^[5, 6]. 然而由于分子结构的多样性及与生物相互作用的复杂性, 正确地确定化合物的毒性作用模式是很困难的. 在数据库分子搜索技术——指纹理论和分子碎片方法基础上发展起来的分子全息 QSAR, 得到了越来越多的应用. 本文试验测

定了一系列苯衍生物对蝌蚪的急性毒性, 并应用分子全息 QSAR 技术发展了具有较高预测能力和稳健性能的结构-活性关系模型.

1 材料与方法

1.1 化合物及受试生物

所测试的苯衍生物都是分析纯, 未进一步提纯. 选用日本长腿蛙 (*Rana japonica*) 的幼体蝌蚪测试生物毒性.

1.2 急性毒性测试方法

日本长腿蛙卵于 2002-04 采自南京市某一蛙

收稿日期: 2004-06-26; 修订日期: 2004-09-05

基金项目: 国家自然科学基金资助项目 (20177008); 上海水产大学博士启动基金 (04-126)

作者简介: 黄宏 (1974 ~), 女, 博士, 讲师, 主要研究方向为有机污染化学及环境毒理学.

塘.所有蛙卵均于产后 2 d 采集并送回实验室,室温下(20℃~25℃)在装有充氧水的 40 cm×40 cm×40 cm 的玻璃槽中喂养.每天喂食 1 次,每 2 d 换水 1 次.毒性测试前 1 d 停止喂食,实验期间也不喂食.选择个体、体重相近、健康的蝌蚪进行毒性试验(试验蝌蚪平均体长 3.0 cm±0.1 cm,体重 0.12 g±0.01 g).在室温和自然光照条件下,将待测化合物用经过 24 h 曝气的自来水配制成不同浓度的 500 mL 溶液.通过预试验确定浓度范围,包括无效应浓度和 100%死亡浓度.每个样品设一个空白,6 个浓度,每个浓度设 3 个重复.每个样品点放入 10 只蝌蚪.每 6 h 更换 1 次受试溶液.12 h 后观察其死亡率,计算 12 h 的半数致死浓度.用每个化合物的半数致死浓度的负对数值(12h-logI/LC₅₀, mol/L)表示化合物的毒性.

1.3 分子模拟与分子全息的产生

分子模拟采用 SYBYL 分子模拟软件^[7],在 SGI O₂ 工作站上进行(IRIX6.5).用 Sketch 选项构建苯及衍生物分子的二维结构.计算部分原子电荷产生一个静电场,用 Gasteiger-Hückel 法计算电荷,采用 SYBYL 中的 HQSAR 模块产生分子全息.分子全息的产生十分类似于 UNITY 指纹的生成.在 UNITY 中,一定大小的碎片可拆分并编成二进制的指纹,然后使用 CRC(cyclic redundancy check)算法将每一相应的碎片映射为 0~2³¹ 内的伪随机整数,这种映射对每个 SLN(SYBYL line notation)串都是唯一的和可重复的^[8].通过这种映射,就可实现“拆分”过程.分子全息的产生方法与上述方法的主要区别在于分子全息并不是象二进制位串那样对每一位点以 0 或 1 的形式编码,而是保留计数方法且以上述每一位点的倍数储存,这样一个分子可表示成一个整形串(见表 1).

上述例子中化合物结构包含有 97(6+18+12+5+14+42)个碎片,它们各自被映射到整形串上.在指纹拆分法中,表征是纯二进制的;而分子全息中,位点能够清楚地说明映射到每一位点上的碎片数目.必须注意的是,当分子全息使用偏最小二乘法分析计算产生一个 QSAR 模型时,一些对特定生物活性贡献较大的分子碎片可能映射到同一位点上,且无法预测,这种现象称为碎片碰撞(fragment collision),这种碰撞可能导致产生的 QSAR 模型的预测能力较差^[9].在 HQSAR 分析中,通过对不同全息长度加以选择,可有效减少碰撞的发生,产生较好的计算模型.

表 1 HQSAR 中分子全息的产生

Table 1 Generation of molecular hologram in HQSAR

Hashed fingerprint	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0
Molecular hologram	0	0	6	18	0	0	0	12	5	14	42	0	0	0	0

1.4 HQSAR 模型的构建与回归分析方法

HQSAR 模型构建需要选择参数:①碎片尺寸,包括 1~1, 3~5, 3~10 和默认的碎片长度 4~7 (一般原子和官能团的比较典型的碎片长度).②全息长度,53~353 之间 11 个缺省的质数.③碎片区分参数包括原子类型(A)、化学键类型(B)和连接性(C)等.根据模型的质量,首先确定较优的碎片尺寸与分子全息长度,然后进一步考虑加入其它碎片区分参数,比如质子给体(D)和氢原子(H),探讨模型质量的变化.采用“Leave-One-Out”交叉验证程序确定最佳主成分数以及最小预测标准误差.分析产生一个“Leave-One-Out”交叉验证的回归系数平方值(r_{cv}^2)和一个标准预测误差(SE_p).采用非交叉验证程序分析产生最终的 HQSAR 模型,用传统的回归系数平方值(r^2)和标准估计误差(SE)来表征模型的拟合程度.

2 结果与讨论

2.1 化合物的毒性数据

试验测定了 46 种苯衍生物对蝌蚪 12 h 暴露急性半致死毒性数据,引用了同实验条件下测试的另外 5 个化合物毒性数据(2-氯苯酚,2,6-二甲基苯酚,2,4-二氯苯酚,2-溴-4-甲基苯酚和 1,3-苯二酚).51 个化合物的名称、化学文摘登记号及半致死毒性(12h-logI/LC₅₀, mol/L)列于表 2 中.试验过程中,所有空白对照均未发现蝌蚪死亡现象.测试化合物的半致死毒性的负对数分布范围从 1,3-苯二酚的 2.07 到 1-溴-2,3-二氯苯的 4.56 之间.

2.2 HQSAR 模型的构建及评价

选择 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 这 11 个缺省值作为分子全息长度,碎片区分参数包括原子类型(A)、化学键类型(B)和连接性(C),最初主成分设定为 10,利用 leave-one-out 交叉验证程序确定最优主成分数,然后根据这个最优主成分数进行 HQSAR 非交叉验证分析,得到模型的相关系数(r^2)和标准误差(SE).不同 HQSAR 模型的分析结果列于表 3.

从数据分析结果可见,以 4~7 为碎片长度得到的模型的分子全息长度为 97,6 个主成分.这个模型

表 2 51 种苯衍生物对蝌蚪的急性致死性毒性
(12h log1/ LC₅₀, mol/ L) 及 HQSAR 模型预测值
Table 2 The acute toxicity (12h log1/ LC₅₀, mol/ L) to
Rana japonica tadpoles and calculated & predicted
by HQSAR model of 51 benzene derivatives

化合物	CAS 登记号	12h log1/ LC ₅₀ ¹⁾ / mol·L ⁻¹		
		实验观测值	HQSAR 拟合值	HQSAR 预测值
1,2,3-三氯苯	87-61-6	4.43	4.49	4.45
1,2,4-三氯苯	120-82-1	4.50	4.44	4.37
1-溴-2,3-二氯苯	56961-77-4	4.56	4.45	4.33
1-溴-2,6-二氯苯	19393-92-1	4.48	4.70	4.64
1,3-二氯苯	541-73-1	3.68	3.84	3.87
1,4-二氯苯	106-46-7	3.85	3.76	3.73
1,2-二氯苯	95-50-1	3.79	3.80	3.80
氯苯	108-90-7	3.20	3.32	3.34
苯酚	108-95-2	2.77	2.75	2.76
2-氯苯酚	95-57-8	3.01 ³⁾	3.10	3.13
4-溴苯酚	106-41-2	3.66	3.55	3.16
4-氯苯酚	106-48-9	3.42	3.13	3.08
4-氟苯酚	371-41-5	2.69	2.85	2.97
2-甲氧基苯酚	1990-5-1	2.65	2.60	2.68
2-甲基苯酚	95-48-7	2.84	2.90	2.92
4-甲氧基苯酚	150-76-5	2.62	2.65	2.79
4-甲基苯酚	106-44-5	3.06	3.02	3.00
4-叔丁基苯酚	831-82-3	4.03	4.00	3.71
2,6-二甲基苯酚	576-26-1	3.32 ³⁾	3.24	3.03
1-萘酚	90-15-3	3.81	3.77	3.80
2-萘酚	135-19-3	3.89	3.90	3.83
2,4-二氯苯酚	121-83-2	3.87 ³⁾	3.68	3.58
2-溴-4-甲基苯酚	6627-55-0	3.72 ³⁾	3.75	3.59
1,3-苯二酚	108-46-3	2.07 ³⁾	2.51	2.71
双酚 A	2)	4.20	4.21	4.03
双酚 E	2)	3.91	3.88	4.18
2,4-二氯苯胺	554-00-7	3.73	3.77	3.82
4-氯苯甲酸	74-11-3	3.42	3.25	3.29
4-溴苯甲酸	586-76-5	3.63	3.44	3.41
水杨酸	69-72-7	2.84	2.73	2.76
5-氯水杨酸	321-14-2	3.01 ³⁾	3.10	3.31
4-羟基苯甲醛	123-08-0	3.08	2.99	3.01
硝基苯	98-95-3	3.29	3.47	3.52
2-硝基甲苯	88-72-2	3.53	3.72	3.78
4-硝基甲苯	99-99-0	3.62	3.62	3.62
2-硝基苯酚	88-75-5	3.50	3.55	3.54
3-硝基苯酚	554-84-7	3.51	3.37	3.33
4-硝基苯酚	123-30-8	3.66	3.38	3.29
1-氯-4-硝基苯	100-00-5	3.93	3.64	3.74
1-溴甲基-4-硝基苯	100-11-8	4.38	4.26	3.92
1-氯甲基-4-硝基苯	100-14-1	4.32	4.37	4.23
4-氯-2-硝基苯酚	89-64-5	3.88	3.98	3.80
2-硝基间苯二酚	601-89-8	3.49	3.61	3.50
2-氯-5-硝基苯胺	6283-25-6	3.47	3.54	3.75
4-硝基-1 萘胺	776-34-1	4.24	4.30	4.22
1,2-二硝基苯	528-29-0	4.05	4.13	4.21
1,3-二硝基苯	99-65-0	4.02	3.99	4.00
2,4-二硝基甲苯	121-14-2	4.06	4.10	4.20
2,4-二硝基苯酚	51-28-5	4.31	4.21	4.08
2,4-二硝基溴苯	584-48-5	4.46	4.56	4.65
2,4-二硝基氯苯	97-00-7	4.34	4.14	4.09

1) 12 h 暴露实验后半致死浓度的负对数 2) 化合物的化学文摘登记号没有提供 3) 毒性数据摘自文献[10]

具有较好的预测能力,而 Leave- One- Out 交叉验证的 r_{cv}^2 是 0.845,交叉验证的标准预测误差为 0.243. 最终的非交叉验证的拟合模型的相关系数 r^2 是 0.933,标准误差为 0.159.接着对 4~7 这个碎片长度范围细分为 4~5、5~6、6~7、6~8、7~8 进行 HQSAR 分析.通过计算得到的最优模型以 6~7 为碎片长度,全息长度为 83,6 个主成分. Leave- One- Out 交叉验证的 r_{cv}^2 是 0.849,交叉验证的标准预测误差为 0.239.最终的非交叉验证拟合模型的相关系数 r^2 为 0.942,标准误差为 0.150.

在前述的 HQSAR 分析中,全息长度是一些缺省的质数,为了提高模型质量,扩大全息长度的搜索范围.设定碎片长度仍为 6~7,全息长度搜索范围为 71~83 和 83~97 之间所有数,结果发现模型质量没有改进.最后又研究了其它的碎片区分常数对模型的影响,结果表明,加入碎片区分常数质子给体 (D) 和氢原子 (H) 进行 HQSAR 分析,反而降低了模型的质量.

通过分析知,对于所研究的 51 种苯衍生物对蝌蚪的急性致死毒性的最佳模型,以 6~7 为碎片尺寸,83 为全息长度,6 个主成分,Leave- One- Out 交叉验证的 r_{cv}^2 是 0.849,交叉验证的标准预测误差为 0.239.最终的非交叉验证的拟合模型的相关系数 r^2 为 0.942,标准误差为 0.150.图 1 分别是用 HQSAR 模型对苯衍生物的急性毒性的模型拟合结果与 Leave- One- Out 交叉验证结果与试验测定值的对照.从图 1 可以看出,用 HQSAR 模型得到的拟合值与预测值和试验结果都非常接近.这些结果说明基于分子全息的 QSAR 模型可以用于预测苯衍生物的生态毒性.

3 结论

试验测定了一系列苯衍生物对蝌蚪的 12 h 急性致死毒性,并应用分子全息方法,发展和建立了 QSAR 模型,对典型的有机污染物苯衍生物的急性毒性进行了预测.探讨了分子全息长度,分子碎片尺寸及区分常数对 HQSAR 模型质量的影响.构建的最优模型以 6~7 为碎片尺寸,83 为全息长度,6 个主成分,Leave- One- Out 交叉验证的 r_{cv}^2 是 0.849,交叉验证的标准预测误差为 0.239;最终的非交叉验证的拟合模型 r^2 为 0.942,标准误差为 0.150.因此模型具有较好的预测能力和稳健性能,可用来对典型污染物——苯衍生物的生态毒性进行预测.

表 3 改变碎片尺寸得到的一系列 HQSAR 模型¹⁾

Table 3 Result of HQSAR for toxicity to guppy of 51 benzene derivatives

碎片尺寸	全息长度	r^2	S. E.	r_{cv}^2	SEp	主成分数	碎片区分	备注
1 ~ 3	61	0.903	0.192	0.840	0.246	6	A,B,C	
3 ~ 5	401	0.907	0.188	0.819	0.263	6	A,B,C	
4 ~ 7 ²⁾	97	0.933	0.159	0.845	0.243	6	A,B,C	
3 ~ 10	353	0.933	0.160	0.739	0.315	6	A,B,C	
4 ~ 5	83	0.910	0.185	0.819	0.262	6	A,B,C	
5 ~ 6	401	0.926	0.166	0.828	0.256	6	A,B,C	
6 ~ 7 ²⁾	83	0.942	0.150	0.849	0.239	6	A,B,C	
6 ~ 8	97	0.937	0.155	0.815	0.266	6	A,B,C	
7 ~ 8	401	0.956	0.130	0.778	0.291	6	A,B,C	
6 ~ 7	83	0.942	0.150	0.849	0.239	6	A,B,C	71 ~ 83 ³⁾
6 ~ 7	83	0.942	0.150	0.849	0.239	6	A,B,C	83 ~ 97 ³⁾
6 ~ 7	257	0.954	0.132	0.804	0.273	6	A,B,C,D	
6 ~ 7	83	0.899	0.194	0.745	0.308	6	A,B,C,H	
6 ~ 7	83	0.899	0.194	0.747	0.307	6	A,B,C,D,H	

1) r_{cv}^2 = leave-one-out 交叉验证的相关系数的平方, SEp = leave-one-out 交叉验证的标准预测误差, 原子类型(A),化学键类型(B),连接性(C), 质子给体(D)和氢原子(H) 2) 最优模型 3) 非缺省的全息长度 .

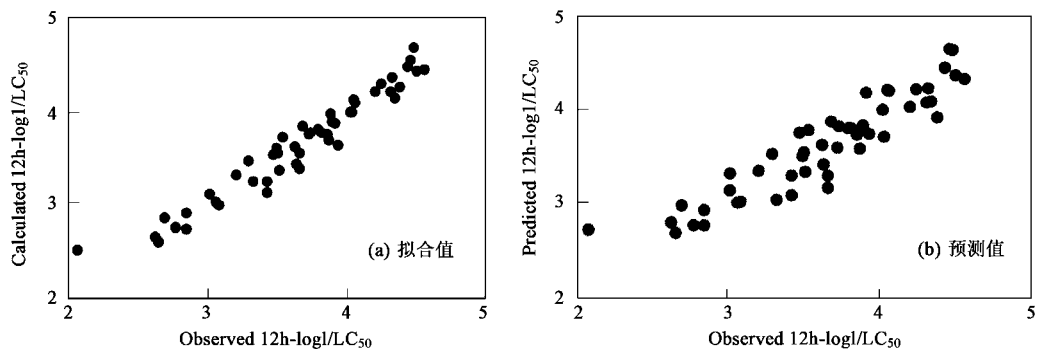


图 1 51 种苯衍生物对蝌蚪急性毒性的 HQSAR 拟合结果和预测结果与实验测定结果的对照

Fig.1 Scatter plot of calculated values (a) and cross-validation predicted values (b) from HQSAR analysis versus observed log 1/ LC₅₀(mol/ L) value of 51 benzene derivatives to *Rana japonica*

参考文献 :

[1] Sixt S, Altschuh J, Brügge mann R. Quantitative structure-toxicity relationships for 80 chlorinated compounds using quantum chemical descriptors [J]. Chemosphere, 1995, 30: 2397 ~ 2414 .

[2] Demichelis SO, Dela Torre FR, Ferrari L, Garcia ME, Salibian A. Tadpoles assay: its application to a water toxicity assessment of a polluted urban river [J]. Environmental Monitoring and Assessment, 2001, 68: 63 ~ 73 .

[3] Mann RM, Bidwell JR. The acute toxicity of agricultural surfactants to the tadpoles of four Australian and two exotic frogs [J]. Environmental Pollution, 2001, 114:195 ~ 205 .

[4] Kurup A, Garg R, Carini DJ, Hanch C. Comparative QSAR: angiotensin II antagonists [J]. Chem. Rev., 2001, 101:2727 ~ 2750 .

[5] Mckim J M, Bradbury S P, Niemi G J. Fish acute toxicity syndromes and their use in the QSAR approach to hazard assessment [J]. Environ. Health Persp., 1987, 71:171 ~ 186 .

[6] Lipnick R L. Base-line toxicity predicted by quantitative structure-activity relationships as a probe for molecular mechanism of toxicity, in: probing bioactive mechanisms Magee P S, Henry D R, Block J H (eds) [M]. ACS symposium series, 1989, 109/ 110:131 ~ 153 .

[7] SYBYL 6.7 user guide [M]. Tripos Inc: St. Louis, USA. 1999 .

[8] Ash S, Cline M, Homer R. SYBYL line notation: a versatile language for chemical structure representation [J]. J. Chem. Inf. Comput. Sci., 1997, 37:71 ~ 79 .

[9] Huang H, Ou WH, Zhao JS, Chen D, Wang LS. A comparative study of quantitative structure-activity relationship methods based on gallic acid derivatives [J]. SAR QSAR Environ. Res., 2004,15:83 ~ 99 .

[10] Wang X D, Dong Y Y, Xu S, Wang L S, Han S K. Acute toxicity of substituted phenols to *Rana japonica* tadpoles and mechanism-based quantitative structure-activity relationship (QSAR) study [J]. Chemosphere, 2001, 44: 447 ~ 455 .