

根据片段常数法估算有机化合物对虹鳟的 LC_{50} 值

习晓环, 徐福留, 曹军, 陶澍* (北京大学城市与环境学系地表过程分析与模拟教育部重点实验室, 北京 100871, E-mail: taos@urban.pku.edu.cn)

摘要: 根据 258 种有机物对虹鳟的 96h 半数致死浓度 (LC_{50}) 实测值建立了预测有机物对虹鳟 LC_{50} 的片段常数估算模型. 讨论了模型的误差和稳健性. 结果表明, 有机物的片段常数及结构因子与虹鳟 $\log 1/LC_{50}$ 之间具有很好的定量关系. 可以根据化合物片段估算其对虹鳟的半数致死浓度. 最终模型的可决系数为 0.9495, 平均误差 0.42 个对数单位. 无论就单一化合物或化合物类别而言, 模型均具有较高的稳健性.

关键词: QSAR; 虹鳟; LC_{50} ; 片段常数

中图分类号: X174 文献标识码: A 文章编号: 0250-3301(2001)04-04-0029

Estimation of LC_{50} of Chemicals to Rainbow Trout by Fragment Constant Method

Xi Xiaohuan, Xu Fuli, Cao Jun, Tao Shu (Dept of Urban and Environ Sci., Peking University, MOE Lab. of Earth Surface Processes, Beijing 100871, China E-mail: taos@urban.pku.edu.cn)

Abstract: A fragment constant model for prediction of 96h LC_{50} of chemicals to rainbow trout was developed based on measured experimental data of 258 chemicals collected from the literature. The accuracy and the robustness of the model were discussed. The coefficient of determination of the model is 0.9495 and the mean residual is 0.42 log-unit. The model is robust for both individual chemical or chemical class.

Key words: QSAR; rainbow trout; fish; LC_{50} ; fragment constant

毒性参数是环境风险评价中不可或缺的. 半数致死浓度 (LC_{50}) 是重要的急性毒性指标之一. 与理化参数相比, 测定毒性参数的过程更复杂, 对所有化合物进行测定的难度极大. 因此, 定量结构-活性关系研究对毒性参数尤其重要.

目前毒理学研究中常用的 QSAR 方法包括辛醇/水分配系数等^[1~4]. 片段常数模型具有误差小、稳健性好的一般优点, 曾被成功地用于预测水-辛醇分配系数 (P_{ow})、土壤吸附系数 (K_{oc}) 和鱼的富集系数 (BCF) 等参数^[5~7]. 本研究在收集各类化合物对虹鳟的 96h 半致死浓度实测资料的基础上, 尝试建立预测 LC_{50} 的片段常数模型.

1 研究方法

研究数据来源于 Tomlin 和 Montgomery 整理的文献^[8,9]. 包括 258 种有机物对虹鳟的

96h 半数致死水相浓度 (LC_{50}), 单位取 $mmol/L$. 文献中仅给出区间数据的化合物, 范围较小者取中位数, 范围较大者 (如 0.2 ~ 80 mg/L) 剔除. 所研究的化合物包括: 氨基甲酸盐、除虫菊酯、氯乙酰、胺类 (苯胺、酰胺、亚胺类)、腈、醇酚醚、酮、N-杂环类、酸、有机磷化合物和其它 (激素、芳香烃衍生物、氧化物及非极性有机物等) 共 11 种. $\log 1/LC_{50}$ 范围跨近 8 个数量级 (-1.107 ~ 6.735).

以 Leo 提出的片段划分原则为基础^[10], 按照朴海善等确定的结构校正原则^[11], 将 258 个化合物划分为 104 种片段和 21 种结构因子.

基金项目: 国家自然科学基金项目 (49971070); 国家杰出青年基金项目 (49525102)

作者简介: 习晓环 (1972), 女, 硕士, 主要研究方向为环境地学.

收稿日期: 2000-09-26

用于 LC_{50} 的片段常数模型的一般形式为: 0.41, 与建模组的 0.40 非常接近.

$$\log \frac{1}{LC_{50}} = \sum_{i=1} n_i f_i + \sum_{j=1} M_j F_j \quad (1)$$

式中, LC_{50} 为有机物对虹鳟的半数致死浓度, f_i 表示第 i 种片段的片段常数, n_i 表示化合物中第 i 种片段的个数, F_j 表示第 j 种结构因子的系数, M_j 则表示第 j 种结构因子的个数. 初步选择的结构因子最终根据逐步回归结果选定. 将数据库中所有 258 种化合物随机分为建模组 (225 种) 和验证组 (33 种). 前者用于回归, 后者用于验证. 验证后利用全部数据建立最终模型. 采用 2 种不同的 jackknife 方法检验模型的稳健性^[12]. 2 种 jackknife 检验分别为 (1) 重复 30 次每次随机抽除 20 个化合物; (2) 依次抽取一类化合物 (共 11 类).

2 结果与讨论

2.1 初步回归模型及其验证

据式(1)以建模组 225 种化合物的片段数和结构特征数为自变量, 以实测 $\log 1/LC_{50}$ 值为因变量进行多元回归. 模型的可决系数为 0.9519, 平均残差 0.40 个对数单位. 图 1 为 225 种化合物的 $\log 1/LC_{50}$ 实测值与计算值的关系, 数据点基本落在 1:1 直线附近, 表现出很好的线性趋势.

对所选择的 21 种结构校正因子的逐步回归结果证实, 这些因子对模型均有贡献, 模型可决系数从没有结构校正的 0.9258 增至 0.9519. 因此, 全部 21 种结构因子均包括在模型中. 将模型应用于检验组的 33 种化合物. 验证组计算值与实测值的关系如图 2 所示.

从图 2 可以看出, 个别化合物表现出相当大的误差. 其原因是验证组中部分化合物具有的特殊片段在建模组中没有出现. 由于分组是随机进行的, 这样的结果不可避免. 验证计算时这些片段常数只能取 0. 其结果是, 它们对 LC_{50} 的贡献没有包括在模型中. 在图 2 中, 这些化合物 (共 6 种) 用实心点表示. 显而易见, 导致个别化合物拟合误差大的原因正在于此. 如果不计这 6 种化合物, 验证组的绝对平均误差为

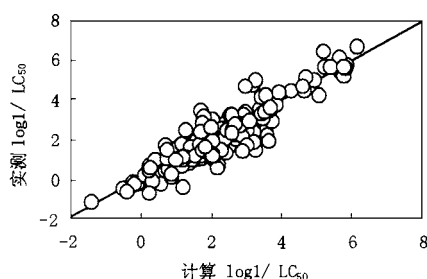


图 1 建模组实测 $\log 1/LC_{50}$ 与初步模型预测 $\log 1/LC_{50}$ 关系

Fig.1 Relationship between the measured and the predicted $\log 1/LC_{50}$ for the modeling set

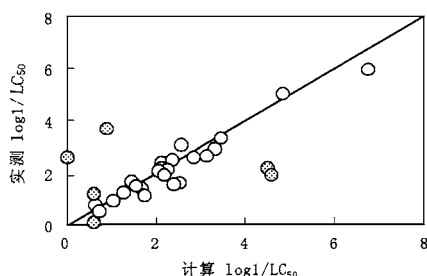


图 2 验证组实测 $\log 1/LC_{50}$ 与初步模型预测 $\log 1/LC_{50}$ 关系

Fig.2 Relationship between the measured and the predicted $\log 1/LC_{50}$ for the validation set

2.2 最终模型

为充分利用信息, 确保模型包含数据库中化合物的所有片段, 在获得满意的验证结果的基础上利用全部数据建立最终模型. 模型的可决系数为 0.9495, 较初步模型略有下降. 全部 258 种化合物的平均绝对误差为 0.42 个对数单位. 误差小于 0.5 个对数单位的化合物占全部化合物的 69.4%. 图 3 显示的最终模型预测结果与实测值关系说明无论从线性关系还是误差分布来看, 模型结果都比较理想. 最终模型获得的有关常数分别在表 1 和表 2 中给出.

根据化合物的亲脂性理论, 亲水性片段的片段常数应为负, 憎水性片段常数值应为正. 但表 1 中多数亲水片段取值为正. 有机化合物致毒作用至少包括吸收与毒作用两个连续过程. 前者与化合物的亲脂性相关, 一般而言, 亲脂性

表 1 最终模型片段常数¹⁾

Table 1 Fragment constants for the final model

片 段	<i>f</i>	<i>f</i> [Ⓛ]	<i>f</i> ^{ⓁⓁ}	片 段	<i>f</i>	<i>f</i> [Ⓛ]	<i>f</i> ^{Ⓛ / Ⓛ}	<i>f</i> ^{ⓁⓁⓁ}
不含 C 和 H				SC(O) N	1. 293			
F	0. 049	0. 007		NC—O	0. 180	0. 358	1. 484	- 1. 327
Cl	0. 269	0. 038		S—C—N—	2. 055			
Br	0. 246	- 0. 335		N—C—S—	1. 597	0. 440		
S	0. 494	0. 600	- 4. 286	OC(O) N	- 0. 439	1. 745		
N	0. 089	0. 237	1. 188	含 C 和 H				
O	0. 074	- 0. 091	0. 166	CH ₃	0. 192	- 0. 455		
NO ₂		0. 533		COOH	- 0. 587	- 0. 999		
SO	- 0. 193			COH	1. 365			
S(O) (O)	0. 232	- 1. 648		CONH	- 0. 237	- 1. 749	1. 128	- 0. 303
S(O) (O) (O)		0. 9098		CONH	1. 325	0. 234	2. 264	- 1. 761
O—P	- 0. 769			CH—NO	- 0. 428	0. 371		0. 831
OP(O) O	3. 063			ON—(CN) CCONHCONH	0. 822			
OP(S) O	0. 350			CH—NOCONH	0. 977			
SP(O) OO	1. 244			N—CH—N—CH—N				1. 992
OP(O) OO	- 1. 089			CONHCONH				- 1. 731
OP(O) SO			2. 584	NHCONO		- 0. 593		
OP(S) OO	1. 628	1. 932		NHCONH		0. 382		
SP(S) OO	2. 348			C—NOCONH	0. 865			
SP(S) O—		3. 173		NC—N(NO ₂) NH	- 0. 442			
SP(O) OS—			2. 044	CHF ₂		0. 425		
含 H 不含 C				NC(O) NH		- 0. 175	- 1. 384	
H	- 0. 018			NHC—N		- 0. 923		
OH	0. 037	- 0. 198		NHCSNH		2. 390		
OP(O) ONH		1. 946		H ₂ NCN	- 0. 330			
OP(S) ONH		0. 055		H ₂ NC(NH) NH	0. 286			
SO ₃ H		- 2. 083		SC—N—OCONH	1. 688			
NHSO ₃		- 0. 366		芳香环或类芳香环上的片段(<i>f</i> ^{AR})				
含 C 不含 H				C—	0. 423			
C	0. 249	0. 204	0. 192	CH	0. 016			
C—N	1. 017	1. 650		C*—	- 0. 251			
C—N—	- 1. 465			N—	0. 791			
CO	- 0. 230	0. 607		N—	- 0. 500			
COO	0. 805	0. 763		NH	- 0. 505			
CF ₃		- 0. 251		C—O	- 0. 699			
C(O) NC(O)	2. 278	- 0. 748		O	0. 368			
C—NO	1. 038	0. 994		S	1. 660			

1) *f*:与脂链相连;*f*[Ⓛ]:从左侧与芳环相连;*f*^{Ⓛ / Ⓛ}:从右侧与芳环相连;*f*^{ⓁⓁ}:与 2 个芳环相连;*f*^{AR}:芳环上片段

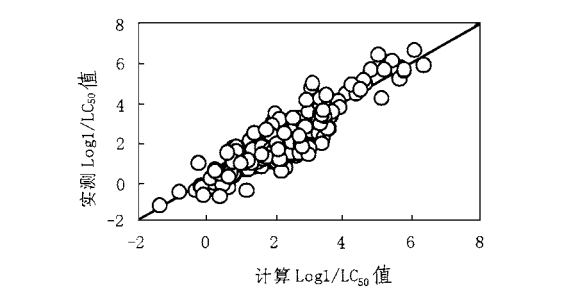


图 3 全部实测 log 1/ LC₅₀与
最终模型预测 log 1/ LC₅₀关系

Fig. 3 Relationship between the measured and the predicted log 1/ LC₅₀ for all chemicals

强的化合物更容易被吸收和富集.但分子量过大时,这样的吸收反而变得困难^[7].毒作用过程则复杂得多,亲脂性强的化合物不一定毒性强.譬

如,癸醇前的烷基醇的毒性随分子增大毒性线性增大,癸醇后随着分子的增大毒性则减小,到十三醇毒性趋近于 0^[13].此外,在片段常数法中,多元回归模型中设计的诸多变量并非正交,特别是片段与结构因子之间以及结构因子与结构因子之间的自相关必然影响回归系数取值.因此,片段常数取值正负也可以是拟合的结果.此外,少数片段在涉及的化合物中出现频率低也是导致个别片段计算结果有偏差的原因.

2. 3 模型稳健性

从数据中重复 30 次随机抽取 20 个化合物进行的 jackknife 检查结果如图 4. 计算 jackknife 可决系数在 0. 9475 到 0. 9579 之间的较小范围内波动,表明就个别化合物而言,模型具有很高的个体稳健性.

表 2 结构因子

Table 2 Structure correction factors

结构因子	常数	结构因子	常数	结构因子	常数	结构因子	常数
F_b	- 0.200	F_{-}	- 0.850	F_{P5}	- 0.078	F_{HP2}	0.294
F_{XCCY}	- 0.449	F_{P0}	0.058	F_{P6}	- 0.142	F_{NH}	0.613
F_{cBr}	0.197	F_{P1}	0.447	F_{mhV}	0.337	F_{NH2}	- 0.483
F_b	- 0.205	F_{P2}	- 0.181	F_{mhGl}	- 0.070		
F_{-}	1.055	F_{P3}	0.223	F_{mhG2}	- 0.071		
$F_{=}$	1.052	F_{P4}	0.035	F_{HPI}	0.656		

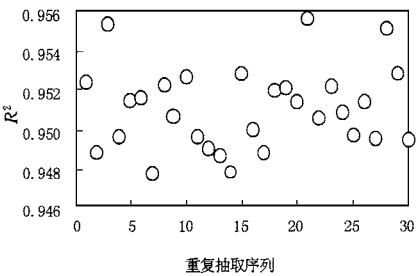
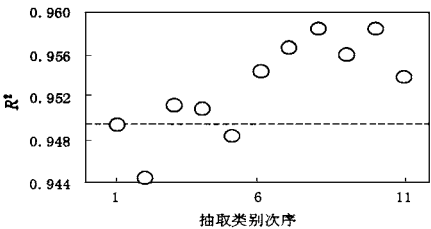


图 4 30 次重复抽取 20 个化合物得到的 jackknife R^2

Fig.4 R^2 from the random set deletion jackknife test 为从化合物类别角度检验模型的稳健性,依次从数据中抽取一类化合物,重复 11 次,得到的 jackknife 可决系数如图 5 所示.多数类别化合物的 jackknife 可决系数都较总体模型可决系数大,尤其是 N 杂环类和有机磷化合物,主要是因为抽取数目较大,且这两类化合物结构复杂.但从总体上看,jackknife 可决系数变化幅度(0.0142)与上述随机抽取 20 个化合物的结果(0.0104)相关不大,可见模型具有相当高的类别稳健性.



从左至右依次为氨基甲酸盐、除虫菊酯、氯乙酰、胺、醇、酚醚、酮与双酮、N 杂环类、酸、有机磷和其它

5 依次抽取一类化合物(共 11 类)得到的 jackknife R^2 Fig.5 R^2 from the class-by-class deletion jackknife test

3 结论

有机化合物的片段常数及结构因子与化合

物对虹鳟的 LC_{50} 的贡献有很好的线性加和关系,可以根据片段常数法估算有机化合物对虹鳟的 LC_{50} .据 258 个实测数据得到的回归模型的可决系数为 0.9495,平均误差 0.42 个对数单位.模型具有相当高的个体和类别稳健性.

参考文献:

1 Nendza, Monika. In Structure- Activity Relationships in Environmental Sciences. Londong: Chapman & Hall, 1998. 73 ~ 77, 155 ~ 185.

2 Kamlet M J, Doherty R M et al. Linear Solvation Energy Relationship. Environ. Sci. Technol., 1988, 22: 503 ~ 509.

3 Hall L H, Kier L B. Estimation of Environmental and Toxicological Properties: Approach and Methodology. Environ. Toxic & Chem., 1989, 8: 19 ~ 24.

4 Nendza, Monika. In Structure- Activity Relationships in Environmental Sciences. London: Chapman & Hall, 1998. 72 ~ 73.

5 Rekker R. The Hydrophobic Fragmental Constant. Elsevier: Amsterdam, 1977.

6 Tao S, Piao HS, Lu XX. Estimation of organic carbon normalized sorption coefficient(K_{oc}) for soils by fragment constant method. Environ. Sci. Technol., 1999, 33: 2719 ~ 2725.

7 Tao S, Hu HY, Lu XX et al. Fragment constant method for predication of fish bioconcentration factors of non-polar chemicals. Chemosphere, 2000, 41: 1563 ~ 1568.

8 Clive Tomlin. The Pesticide Manual: incorporating the Agrochemicals Handbook(10th ed). Farnham, Surry, British Crop Protection Council, 1994.

9 John H Montgomery. Agrochemicals desk reference: environmental data. Boca Raton: Lewis Pub, 1993.

10 Leo A J, Lyman, W J, Rosenblatt D H Ed, McGraw- Hill. In: Handbook of Chemical Property Estimation Methods. New York: 1982. 1 ~ 54.

11 朴海善, 陶澍, 胡海瑛等. 估算有机化合物土壤吸着系数 K_{oc} 的片段常数法. 土壤学报, 2000, 37: 209 ~ 216.

陶澍, 卢晓霞, 曹军, 胡海瑛. 有机化合物 K_{oc} 分子拓扑极性校正模型的稳健性检验. 土壤学报, 2000, 37: 34 ~ 40.

13 Veith G D, Call D J et al. Structure- Activity Relationships for the Fathead Minnow, Pimephales Promelas: Narcotic Industrial. Can. J. Fish. Aquat. Sci., 1983, 40: 743 ~ 748.