应用遗传算法建立云杉针叶表面 PCDD/ Fs 光解半衰期的预测模型

牛军峰 ,余刚 * ,韩文亚
(清华大学环境科学与工程系持久性有机污染物研究中心 ,北京 100084)

摘要:采用量子化学 PM3 算法计算 PCDD/ Fs 的量子化学参数 ,应用遗传算法(Genetic Algorithm , GA)对所建模型中的变量进行筛选 ,建立了能预测吸附到云杉[*Picea abies* (L.) Karst.]针叶表面的 PCDD/ Fs 光解半衰期($t_{1/2}$)的定量结构-性质关系(QSPR)模型 ,认为影响 PCDD/ Fs 光解速率的主要因素是分子最高占据轨道能(E_{HOMO})、前线分子轨道能量差(E_{LUMO} - E_{HOMO})和平均极化率(α). 随着 E_{HOMO}和 α 的增大 ;其 lg $t_{1/2}$值增大 ,PCDD/ Fs 的 lg $t_{1/2}$值与其(E_{LUMO} - E_{HOMO})值呈抛物线关系 ,当 E_{LUMO} - E_{HOMO} = 7. 847 时 ,lg $t_{1/2}$有最小值 ;当 E_{LUMO} - E_{HOMO} ≤7. 847 时 ,lg $t_{1/2}$值随着 E_{LUMO} - E_{HOMO}的增大而减小 ;当 E_{LUMO} - E_{HOMO} ≥7. 847 时 ,lg $t_{1/2}$值随着 E_{LUMO} - E_{HOMO}的增大而增大.

关键词:遗传算法 ;PCDD/ Fs ;定量结构-性质关系(QSPR) ;量子化学参数
中图分类号:X173 **文献标识码**:A **文章编号**:0250-3301(2005)02-0028-06

Prediction of Photolysis Half-Lives of PCDD / Fs Adsorbed on Spruce Needles Optimized by Genetic Algorithm

NIU Jun-feng , YU Gang , HAN Wen-ya
(Research Center for Persistent Organic Pollutants , Department of Environmental Science and Engineering , Tsinghua University , Beijing 100084 , China)

Abstract :Adopting quantum chemical parameters of PCDD/ Fs computed with quantum chemical PM3 algorithm , quantitative structure-property relationship (QSPR) model , which could predict photolysis half-life ($t_{1/2}$) of PCDD/ Fs adsorbed to spruce [*Picea abies* (L.) Karst.] needle surfaces , is established using genetic algorithm (GA) algorithm. It is considered that the main factors affecting lg $t_{1/2}$ of PCDD/ Fs are the energy of the highest occupied molecular orbital (E_{HOMO}) , E_{LUMO} - E_{HOMO} and average molecular polarizability (α). The lg $t_{1/2}$ values increase with the increasing of E_{HOMO} and α. The relationship between the lg $t_{1/2}$ values and E_{LUMO} - E_{HOMO} is a parabolic curve. The lg $t_{1/2}$ values increase with the increasing of E_{LUMO} - E_{HOMO} when E_{LUMO} - E_{HOMO} ≥7. 847 and decrease with the increasing of when E_{LUMO} - E_{HOMO} ≤7. 847.

Key words :genetic algorithm ; PCDD/ Fs ; QSPR ; quantum chemical descriptors

多氯代二苯并-对-二[illegible]和多氯代二苯并呋喃(PCDD/ Fs) 是环境中广泛存在的典型持久性有机污染物(POPs)[1,2] ,城市生活垃圾焚烧、含铅汽油、煤炭、木材和泥炭的燃烧、造纸过程中的氯气漂白、冶金过程等均能产生 PCDD/ Fs[1~3]. PCDD/ Fs 类有机污染物对生物具有致癌性、致畸性、生殖毒性和免疫毒性[4,5] ;近年来的研究还表明 ,当它们在环境中长期低剂量存在时 ,可使生物体的内分泌功能发生紊乱[6] ,引起了国内外广泛关注[4~8].

一些研究表明 ,植物表面光解过程对于有机污染物从大气到食物链的迁移有重要的影响[8]. PCDD/ Fs 在植物表面上的光解主要发生在植物叶面角质层中[8,9] ,其中 ,光解半衰期($t_{1/2}$) 对于考察此类污染物的环境行为并进行环境风险评价有重要的作用. 由于实验测定 PCDD/ Fs 的 $t_{1/2}$需要特殊设备 ,消耗大量的时间和财力 ,因此发展 PCDD/ Fs 光解行为的定量结构-性质关系(quantinative structure-property relationship , QSPR)[10~12]就显得尤为必要. 在进行 QSPR 的研究中 ,对于变量的取舍是建模的关键. 对于变量较多的 QSPR 研究 ,变量的取舍通常依赖于逐步回归分析. 然而 ,逐步回归分析有其自身的不足之处 ,例如在一些情况下 ,正向与反向逐步多元回归分析所建立的模型有可能不同[12] ,从而影响相关机理的解释. 本研究以容易计算得到且具有明确物理-化学意义的分子量子化学参数为自变量 ,采用在化学计量学方面具有广泛应用前景的遗传算法(Genetic Algorithm , GA) 对模型中的变

收稿日期:2004-03-23 ;**修订日期**:2004-04-28
基金项目:国家自然科学基金资助项目(50238020) ;中国博士后科学基金资助项目[中博基(2003) 03]
作者简介:牛军峰(1974～) ,男 ,河南开封人 ,博士 ,主要研究方向为持久性有机污染物(POPs)的环境行为及其治理技术.
* 联系人

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

量进行筛选,建立了具有较强预测能力并有助于对 PCDD/ Fs 光解机理解释的 QSPR 模型.

1　材料与方法

1.1　GA 方法原理

GA 是一种借鉴生物界自然选择和进化机制发展起来的高度并行、随机、自适应搜索算法[13],已获得广泛应用. GA 克服了传统优化方法容易陷入局部极值的特点,是一种全局优化算法[13,14]. 遗传算法的具体实现步骤如下:

(1) 随机或根据一定的限制条件产生一个给定大小的初始解("种群"),每个个体用一个向量 x 进行编码,称为一条"染色体",向量的分量代表基因,它对应个体的某一决策变量.

(2) 遗传算法根据适应度函数来评价个体的优劣,作为以后遗传操作的依据. 计算群体中每条染色体 $x_i(i=1,2,3,\cdots,n)$ 所对应的目标函数值,并以此计算"适值" F_i,按照 F_i 的大小来评价该个体的好坏.

(3) 根据适值的大小,以优胜劣汰的机制,将适值差的染色体淘汰掉,对幸存的染色体根据其适值的好坏,按一定概率随机选择配对,然后 2 个匹配的个体位串根据随机选取的交叉点进行交叉繁殖,产生一队新的个体串.

(4) 通过交叉和变异运算,产生子代. 交叉是最主要的遗传运算,它同时对 2 个染色体操作,组合两者的特征产生新的后代. 交叉的最简单方法是在双亲(2 个父辈的个体)的染色体上随机地选择一个断点,将断点的右段相互交换,从而形成 2 个新的后代. 变异则是一种基本运算,它在染色体上自发地产生随机的变化. 一种简单的变异方法是替换一个或多个基因,在遗传算法中,变异可以提供初始种群中不含有的基因,或找到选择过程中丢失的基因,为种群提供新的内容,拓展搜索空间.

这样,经过若干代之后,算法收敛于最优种群,其中很可能存在问题的最优解或次优解.

1.2　研究对象

研究对象为 70 个 PCDD/ Fs 分子,数据取自 2001 年 7 月德国慕尼黑(纬度 48.2 °N) 夏季晴空下云杉(*Picea abies* (L.) Karst.)针叶表面吸附的 PCDD/ Fs 的光解半衰期($t_{1/2}$)实验测定值. 太阳光暴露时间从每天的 09:00 到 18:00[15]. 实验数据如表 1 所示.

1.3　参数的计算与选取

采用 MOPAC(6.0 版)软件中的 PM3 算法计算这些化合物的量子化学参数. 控制 MOPAC 计算的关键词为: PM3、ESP、POLAR、BONDS、ENPART、DIPOLE、PRECISE、NOINTER. 共选取了 14 个量子化学参数. 这些量子化学参数是主要反映分子整体特征的分子结构参数,如相对分子质量(M_r)、分子生成热(ΔH_f)、平均分子极化率(α)和偶极矩(μ)、分子最高占据轨道能(E_{HOMO})、分子最低未占据轨道能(E_{LUMO})、分子总的核核排斥能(CCR)、电子能(EE)、分子总能量(TE)、PCDD/ Fs 分子中 C 原子所带最大负电荷(q_C^-)、H 原子所带最大正电荷(q_H^+)、O 原子所带最大负电荷(q_O^-)、PCDD/ Fs 分子中 Cl 原子所带最大正电荷(q_{Cl})以及和这个 Cl 相连的 C 原子所带的最大负电荷($q_{Cl\text{-}C}^-$)等. 其中,极化率和偶极矩单位为原子单位(a.u.);能量单位为电子伏特(eV);电荷单位为原子电荷单位(a.c.u.). 对有机物分子光解 QSPR[10,11]的研究表明,分子的前线轨道能 $E_{LUMO}-E_{HOMO}$、$(E_{LUMO}-E_{HOMO})^2$ 和 $E_{LUMO}+E_{HOMO}$ 对于模型的建立也有着重要的影响. 因此将这 3 个参数也作为自变量进行考察. 其中,$E_{LUMO}-E_{HOMO}$ 和 $E_{LUMO}+E_{HOMO}$ 分别和分子的绝对硬度和电负性有关.

1.4　统计分析

采用 STATGRAPHICS(4.0 版)软件[16]进行逐步回归分析,遗传算法用 MATLAB 语言编程,初始种群数目为 70,交叉概率为 0.3,变异概率为 0.1,复制策略采用最优保留方法. 采用复相关系数(R)、经自由度校正的复相关系数(R_{adj})、拟合值的标准误差(SE)、方差分析的方差比(F)和回归的显著性水平(p)以表征模型的优劣.

2　结果与讨论

2.1　逐步回归模型

以实验得到的 70 个 PCDD/ Fs 的 $\lg t_{1/2}$ 值(表 1)为因变量,以计算得到的 17 种量子化学参数为自变量分别进行正向和反向逐步回归分析,模型拟合分别得到了逐步回归分析的方程.

正向逐步回归分析方程:

$$\lg t_{1/2} = (1.281 \pm 0.075) + (3.012 \pm 0.490) \times 10^{-5}\mathrm{CCR} - (2.144 \pm 0.490) \times 10^{-3}\Delta H_f \quad (1)$$

$n=70$, $R^2=0.655$, $R^2_{adj}=0.645$, SE = 0.064, $F=63.70$, $p=0.0000$

式中,n 为回归分析的样品数.

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 1 云杉针叶表面 PCDD/ Fs 的光解半衰期 $t_{1/2}$/ h

Table 1 Photolysis half-lives ($t_{1/2}$) of the PCDD/ Fs on spruce needle surfaces/ h

No.	PCDD/ Fs	实验值[1)]	逐步回归分析		遗传算法	
		lg $t_{1/2}$	lg $t_{1/2}$(预测值)	残差[2)]	lg $t_{1/2}$(预测值)	残差[2)]
1	2,3,7,8-TCDD	1.74	1.83	-0.09	1.80	-0.06
2	1,3,6,8-TCDD	1.90	1.84	0.06	1.85	0.05
3	1,3,7,9-TCDD	1.90	1.85	0.05	1.83	0.07
4	1,3,7,8-TCDD	1.94	1.83	0.11	1.81	0.13
5	1,2,6,8-TCDD	1.91	1.80	0.11	1.84	0.07
6	1,4,7,8-TCDD	1.76	1.81	-0.05	1.84	-0.08
7	1,2,3,9-TCDD	1.90	1.80	0.10	1.82	0.08
8	1,2,6,9-TCDD	1.79	1.80	-0.01	1.92	-0.13
9	1,2,6,7-TCDD	1.89	1.80	0.09	1.86	0.03
10	1,2,8,9-TCDD	1.95	1.82	0.13	1.85	0.10
11	1,2,3,7,8-PeCDD	1.76	1.87	-0.11	1.85	-0.09
12	1,2,3,6,8-PeCDD	1.74	1.85	-0.11	1.86	-0.12
13	1,2,4,7,8-PeCDD	1.91	1.86	0.05	1.86	0.05
14	1,2,3,7,9-PeCDD	1.82	1.87	-0.05	1.85	-0.03
15	1,2,3,6,9-PeCDD	1.89	1.85	0.04	1.88	0.01
16	1,2,4,6,7-PeCDD	1.85	1.86	-0.01	1.90	-0.05
17	1,2,4,8,9-PeCDD	1.84	1.86	-0.02	1.88	-0.04
18	1,2,3,4,6-PeCDD	1.82	1.87	-0.05	1.87	-0.05
19	1,2,3,6,7-PeCDD	1.82	1.86	-0.04	1.86	-0.04
20	1,2,3,8,9-PeCDD	1.85	1.87	-0.02	1.85	0.00
21	1,2,3,4,7,8-HxCDD	1.96	1.95	0.01	1.91	0.05
22	1,2,3,6,7,8-HxCDD	1.88	1.93	-0.05	1.90	-0.02
23	1,2,3,7,8,9-HxCDD	1.97	1.94	0.03	1.90	0.07
24	1,2,3,4,6,9-HxCDD	1.87	1.93	-0.06	1.93	-0.06
25	1,2,3,4,6,7-HxCDD	1.97	1.95	0.02	1.91	0.06
26	1,2,3,4,6,7,8-HpCDD	1.92	2.00	-0.08	1.96	-0.04
27	1,2,3,4,6,7,9-HpCDD	1.98	1.99	-0.01	1.96	0.02
28	OCDD	2.02	2.08	-0.06	2.02	0.00
29	2,3,7,8-TCDF	1.66	1.70	-0.04	1.70	-0.04
30	1,3,6,8-TCDF	1.76	1.72	0.04	1.68	0.08
31	1,4,6,8-TCDF	1.70	1.68	0.02	1.66	0.04
32	1,3,4,8-TCDF	1.64	1.68	-0.04	1.67	-0.03
33	1,2,7,8-TCDF	1.65	1.70	-0.05	1.67	-0.02
34	1,2,6,7-TCDF	1.73	1.68	0.05	1.68	0.05
35	2,4,6,7-TCDF	1.72	1.67	0.05	1.69	0.03
36	1,2,6,9-TCDF	1.61	1.67	-0.06	1.64	-0.03
37	2,3,4,6-TCDF	1.64	1.68	-0.04	1.69	-0.05
38	2,3,6,7-TCDF	1.70	1.68	0.02	1.70	0.00
39	3,4,6,7-TCDF	1.72	1.67	0.05	1.68	0.04
40	2,3,4,7,8-PeCDF	1.74	1.77	-0.03	1.76	-0.02
41	1,3,6,7,8-PeCDF	1.68	1.77	-0.09	1.74	-0.06
42	1,3,4,7,8-PeCDF	1.70	1.76	-0.06	1.77	-0.07
43	1,2,4,7,8-PeCDF	1.69	1.75	-0.06	1.74	-0.05
44	1,2,4,6,7-PeCDF	1.69	1.74	-0.05	1.72	-0.03
45	1,3,4,6,9-PeCDF	1.65	1.73	-0.08	1.71	-0.06
46	1,2,3,4,6-PeCDF	1.67	1.73	-0.06	1.72	-0.05
47	1,2,3,7,9-PeCDF	1.65	1.74	-0.09	1.74	-0.09
48	1,2,3,6,7-PeCDF	1.74	1.72	0.02	1.74	0.00
49	1,2,6,7,9-PeCDF	1.72	1.74	-0.02	1.72	0.00
50	1,2,3,6,9-PeCDF	1.82	1.72	0.10	1.72	0.10
51	2,3,4,6,8-PeCDF	1.72	1.74	-0.02	1.74	-0.02
52	1,2,3,4,9-PeCDF	1.71	1.74	-0.03	1.71	0.00
53	1,2,4,8,9-PeCDF	1.67	1.74	-0.07	1.71	-0.04

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

续表 1

No.	PCDD/ Fs	实验值[1)]	逐步回归分析		遗传算法	
		lg $t_{1/2}$	lg $t_{1/2}$(预测值)	残差[2)]	lg $t_{1/2}$(预测值)	残差[2)]
54	1,2,3,8,9-PeCDF	1.71	1.74	-0.03	1.71	0.00
55	2,3,4,6,7-PeCDF	1.78	1.75	0.03	1.75	0.03
56	1,2,3,6,7,8-HxCDF	1.87	1.80	0.07	1.81	0.06
57	1,2,3,7,8,9-HxCDF	1.90	1.81	0.09	1.81	0.09
58	2,3,4,6,7,8-HxCDF	1.85	1.82	0.03	1.81	0.04
59	1,2,3,4,6,8-HxCDF	1.78	1.79	-0.01	1.80	-0.02
60	1,2,4,6,7,8-HxCDF	1.76	1.82	-0.06	1.80	-0.04
61	1,2,4,6,7,9-HxCDF	1.81	1.81	0.00	1.79	0.02
62	1,2,4,6,8,9-HxCDF	1.92	1.81	0.11	1.79	0.13
63	1,2,3,4,6,7-HxCDF	1.76	1.80	-0.04	1.82	-0.06
64	1,2,3,6,7,9-HxCDF	1.85	1.79	0.06	1.84	0.01
65	1,2,3,4,8,9-HxCDF	1.79	1.81	-0.02	1.79	0.00
66	1,2,3,4,6,7,8-HpCDF	1.93	1.87	0.06	1.92	0.01
67	1,2,3,4,7,8,9-HpCDF	1.93	1.86	0.07	1.92	0.01
68	1,2,3,4,6,7,9-HpCDF	1.89	1.86	0.03	1.92	-0.03
69	1,2,3,4,6,8,9-HpCDF	1.93	1.86	0.07	1.89	0.04
70	OCDF	2.00	1.94	0.06	2.00	0.00

1) 数据取自文献[15]　2) 残差 = lg $t_{1/2}$(实验值) - lg $t_{1/2}$(预测值).

反向逐步回归分析方程:

$$\lg t_{1/2} = (1.154 \pm 0.143) + (2.173 \pm 0.212) \times 10^{-3} M_r + (1.819 \pm 0.681) q_C^- - (2.370 \pm 0.612) q_O^- \quad (2)$$

$n = 70$, $R^2 = 0.667$, $R^2_{adj} = 0.652$, SE = 0.063, $F = 44.13$, $p = 0.0000$

从建立的方程可以看出,正向逐步回归分析和反向逐步回归分析分别得到了不同的方程,不同的变量分别被选入了这 2 个方程中,这就影响了对于 PCDD/ Fs 光解机理的解释. 表 1 中以反向逐步回归分析方程为例列出了 PCDD/ Fs 的 lg $t_{1/2}$的预测值.

2.2　GA 模型

遗传算法控制参数中,初始种群数目为 70,交叉概率为 0.3,变异概率为 0.1,遗传迭代次数为 200. 本文采用因变量的预测值和实验值的复相关系数为适应度函数. 通过 GA 进行变量筛选得到了 E_{HOMO}、前线分子轨道能量差 $E_{LUMO} - E_{HOMO}$以及平均极化率 α等参数对 PCDD/ Fs 的光解具有重要的影响. 因此,以实验得到的 70 个 PCDD/ Fs 的 lg $t_{1/2}$值(表 1)为因变量,以通过 GA 筛选得到的量子化学参数为自变量进行拟合,其模型拟合结果见表 2. 根据表 2,可得到如下的 QSPR 方程:

$$\lg t_{1/2} = (7.299 \pm 1.786) \times 10 + (2.732 \pm 1.136) \times 10^{-1} E_{HOMO} + (7.087 \pm 0.694) \times 10^{-3} \alpha - (1.783 \pm 0.448) \times 10 (E_{LUMO} - E_{HOMO}) + (1.136 \pm 0.281) (E_{LUMO} - E_{HOMO})^2 \quad (3)$$

$n = 70$, $R^2 = 0.721$, $R^2_{adj} = 0.704$, SE = 0.058, $F = 42.00$, $p = 0.0000$

表 2　GA 拟合结果

Table 2　The statistical results by GA

自变量	系数	SE	t-统计值	显著性水平(p)
常数	(7.299) ×10	1.786 ×10	4.087	0.000
E_{HOMO}	2.732×10^{-1}	1.136×10^{-1}	2.404	0.019
α	7.087×10^{-3}	0.694×10^{-3}	10.204	0.000
$E_{LUMO} - E_{HOMO}$	1.783 ×10	0.448 ×10	-3.983	0.000
$(E_{LUMO} - E_{HOMO})^2$	1.136	0.281	4.050	0.000

方程(3)的 F 和 p 值表明这个 QSPR 方程的相关系数是显著的. 由于方程(3)具有较大的相关系数和较小的预测值标准误差,且方程的预测值与实验值接近(表 1 和图 1),因此方程(3)可以应用于云杉

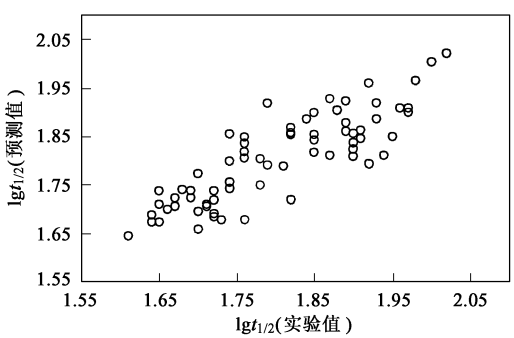

图 1　方程(3)中 lg $t_{1/2}$的实验值与预测值之间的相关关系

Fig. 1　Plot of measured and predicted lg $t_{1/2}$ for PCDD/ Fs

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

针叶表面 PCDD/ Fs 光解半衰期的预测. 从表 2 所列的 t-统计值的大小和符号及 p 值可以看出各个变量的相对重要性.

方程(3)表明, PCDD/ Fs 的前线分子轨道能 E_{HOMO}、前线分子轨道能量差 E_{LUMO} - E_{HOMO}以及平均极化率 α 对于 PCDD/ Fs 的光解起着重要的作用,这与以前不同介质的研究相一致[10,11]. 从方程(3)可以得到如下结论: ①lg $t_{1/2}$与 E_{HOMO}呈正相关,即 PCDD/ Fs 的光解半衰期随着 E_{HOMO}的增大而增大. E_{HOMO}可以表示为化合物提供电子的能力,这表明,云杉针叶表面 PCDD/ Fs 分子越容易提供电子,其光解能力越弱. 这与以前对于水相中 PCDDs 光解的 QSPR 研究结果不一致[10],可能是不同的环境介质中 PCDD/ Fs 的光解机理有所不同[11]. ②随着 α 增大, PCDD/ Fs 相应的光解半衰期增大. 研究表明,还原脱氯对于云杉针叶表面 PCDD/ Fs 的光化学降解起着重要的作用[15]. 其中, α 表示分子的平均极化率,它与分子体积成正比,因此, PCDD/ Fs 的光解速率随着氯原子取代数目的增加而降低,这可能是由于氯原子取代数目越多,"空间位阻效应"越大,进而导致了高氯代 PCDD/ Fs 难以发生降解. 杨曦等[17]通过应用半经验量子化学计算方法研究发现, UV-A(315～400 nm)和 UV-B(280～315 nm)辐射下,2,3,7,8-TCDD 的光解反应以 C—O 键断裂为主,而 UV-C(200～280 nm)辐射下脱氯产物增多. 本研究中太阳光的 UV 波段主要是 UV-A 和 UV-B,而在光解中间过程中检测到了低氯代 PCDD/ Fs 增多的现象[15],这可能是 PCDD/ Fs 存在的介质对光解途径有影响. ③PCDD/ Fs 的光解半衰期与其前线分子轨道能量差 E_{LUMO} - E_{HOMO}呈抛物线关系,当 E_{LUMO} - E_{HOMO} = 7.847 时, lg $t_{1/2}$有最小值;当 E_{LUMO} - E_{HOMO} ≤7.847 时, PCDD/ Fs 的光解半衰期 lg $t_{1/2}$随着 E_{LUMO} - E_{HOMO}的增大而减小;当 E_{LUMO} - E_{HOMO} ≥7.847 时, PCDD/ Fs 的光解半衰期 lg $t_{1/2}$随着 E_{LUMO} - E_{HOMO}的增大而增大;对于多环芳烃(PAHs)的光解速率常数的 QSPR 研究也得到了类似的结论[12]. 从以上的结果可以看出,模型(3)有助于对云杉针叶表面 PCDD/ Fs 光解机理的进一步解释.

2.3 模型的评价和验证

为了考察所建立模型(3)的预测能力,在表 1 中,将样品数分为训练集和预测集,以第 1 号样品开始,每间隔 9 个样品即第 11 号样品、第 21 号样品、第 31 号样品……,依此类推. 每次共有 7 个样品从总的样品中被剔除并作为预测集,其余的作为训练集建立模型并对预测集进行预测,直到所有的样品都被预测一次. 在训练集中,模型中样品的 R 值在 0.830～0.870 之间, SE < 0.063, p < 0.0001;在预测集中,将样品的预测值对实验值作图,如图 2 所示. 模型的预测值和实验值的 R = 0.802, SE = 0.065, p = 0.0000. 从结果可以看出,预测值和实验值有较高的相关系数和小的 SE 值,这表明所建立的预测模型稳定可靠.

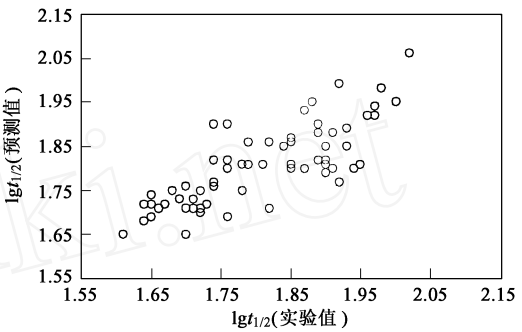

图 2 模型评价中 lg $t_{1/2}$的实验值与预测值之间的相关关系

Fig. 2 Plot of measured and predicted lg $t_{1/2}$ for PCDD/ Fs in model evaluation

在所建立的模型中,各自变量之间的共线性是衡量方程可靠性的重要指标之一,共线性通常用变量间的自相关系数矩阵表示,方程(3)的检验结果列于表 3,由表 3 可以看出, E_{HOMO}和 E_{LUMO} - E_{HOMO}之间简单相关系数为 0.580;在 p ≥0.001 的水平上,相关关系显著. 但表 3 中相关系数的值都较小. 可认为方程(3)中参数间无明显的共线性关系,本研究根据 GA 得到的 QSPR 方程对于 PCDD/ Fs 光解机理的解释有一定的意义.

表 3 部分分子结构参数的相关系数

Table 3 Correlation coefficients between some quantum chemical descriptors (p < 0.05)

	E_{HOMO}	α	E_{LUMO} - E_{HOMO}
E_{HOMO}	1.000		
α	- 0.186	1.000	
E_{LUMO} - E_{HOMO}	0.580	- 0.508	1.000

由方程(1)～(3)可以看出,从遗传算法得到的模型(3)的 R^2 = 0.721 高于正向或反向逐步回归分析得到的相应模型的 R^2,遗传算法得到的预测值的 SE 则低于逐步回归分析得到的相应模型的 SE,因此,用遗传算法得到的模型较常规逐步回归分析得到的相应模型的预测能力有优势,能较好地预测

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

PCDD/ Fs 在云杉针叶表面光解的 lg $t_{1/2}$值. 结果表明,通过 GA 进行变量选取,可优化校正模型,使其具有较强的预测能力,得到的模型也易于对光解机理的解释.

3　结论

应用 GA 对 QSPR 研究中的变量进行筛选,建立了能预测吸附到云杉针叶表面的 PCDD/ Fs 光解半衰期的 QSPR 模型. 认为影响 PCDD/ Fs 光解速率的主要因素是分子最高占据轨道能(E_{HOMO})、前线分子轨道能量差(E_{LUMO} - E_{HOMO})和平均极化率(α),随着 E_{HOMO}和 α 的增大,其 lg $t_{1/2}$值增大,PCDD/ Fs 的 lg $t_{1/2}$值与其(E_{LUMO} - E_{HOMO})值呈抛物线关系,当 E_{LUMO} - E_{HOMO} = 7.847 时,lg $t_{1/2}$有最小值;当 E_{LUMO} - E_{HOMO} ≤7.847 时,lg $t_{1/2}$值随着 E_{LUMO} - E_{HOMO}的增大而减小;当 E_{LUMO} - E_{HOMO} ≥7.847 时,lg $t_{1/2}$值随着 E_{LUMO} - E_{HOMO}的增大而增大. 利用 GA 对 QSPR 研究中的变量进行筛选,得到的结果较传统的逐步多元回归分析更准确,模型有更强的预测能力并易于对 PCDD/ Fs 光解机理的解释.

参考文献:

[1] Cheng P-S, Hsu M-S, Ma E, Chou U, Ling Y-C. Levels of PCDD/ Fs in ambient air and soil in the vicinity of a municipal solid waste incinerator in Hsinchu[J]. Chemosphere, 2003, **52**: 1389～1396.

[2] Hoekstra E J, deWeerd H, de Leer E W B, Brinkman U A T. Natural formation of chlorinated phenols, dibenzo-p-dioxins, and dibenzofurans in soil of a Douglas fir forest[J]. Environ. Sci. Technol., 1999, **33**: 2543～2549.

[3] Meharg A A, Osborn D. Dioxins released from chemical accidents[J]. Nature, 1995, **375**: 353～354.

[4] Boening D W. Toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin to several ecolgical receptor groups: a short review[J]. Ecotoxicol. Environ. Safety, 1998, **39**: 155～163.

[5] Kaiser J. Just how bad is dioxin? [J] Science, 2000, **288**: 1941～1944.

[6] Younes M. Specific issues in health risk assessment of endocrine disrupting chemicals and international activities[J]. Chemosphere, 1999, **39**: 1253～1257.

[7] Bunge M, Adrian L, Kraus A, Opel M, Lorenz W G, Andreesen J R, Görisch H, Lechner U. Reductive dehalgenation of chlorinated dioxins by an anaerobic bacterium[J]. Nature, 2003, **421**: 357～360.

[8] Barber J L, Thomas G O, Kerstiens G, Jones K C. Current issues and uncertainties in the measurement and modeling of air-vegetation exchange and within-plant processing of POPs[J]. Environ. Pollut., 2004, **128**: 99～138.

[9] Schuler F, Schmid P, Schlatter C H. Photodegradation of polychlorinated dibenzo-*p*-dioxins and dibenzofuzans in cuticular waxes of laurel cherry (*Prunus laurocerasus*)[J]. Chemosphere, 1998, **36**: 21～34.

[10] Chen J W, Quan X, Peijnenburg W J G M, Yang F L. Quantitative structure-property relationships (QSPRs) on direct photolysis quantum yields of PCDDs[J]. Chemoshpere, 2001, **43**: 235～241.

[11] Chen J W, Quan X, Yang F L, Peijnenburg W J G M. Quantitative structure-property relationships on photodegradation of PCDD/ Fs in cuticular waxes of laurel cherry (*Prunus laurocerasus*)[J]. Sci. Total Environ., 2001, **269**: 163～170.

[12] 陈景文. 有机污染物定量结构-性质关系与定量结构-活性关系[M]. 大连:大连理工大学出版社, 1999.

[13] Holland J H. Adaptation in natural and artificial systems[M]. MI: University of Michigan Press, 1975.

[14] Wegner J K, Zell A. Prediction of Aqueous Solubility and Partition Coefficient Optimized by a Genetic Algorithm Based Descriptor Selection Method[J]. J. Chem. Inf. Comput. Sci., 2003, **43**: 1077～1084.

[15] Niu J F, Chen J W, Henkelmann B, Quan X, Yang F L, Kettrup A, Schramm K-W. Photodegradation of PCDD/ Fs adsorbed on spruce (*Picea abies* (L.) Karst.) needles under sunlight irradiation[J]. Chemosphere, 2003, **50**: 1217～1225.

[16] STSC. Inc. and Statistical Graphics Corporation. STATGRAPHIC, Version 4.0, 1985.

[17] 杨曦, 余刚, 王连生. 2,3,7,8-四氯代二苯并[illegible]光解行为的量子化学[J]. 科学通报, 2002, **47**(4): 269～274.

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net