The use of fisher's discriminant analysis in the development of classification models for the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill

NIU JunFeng and YU Gang*

(Department of Environmental Science and Engineering, Research Centre for Persistent Organic Pollutants, Tsinghua University, Beijing, 100084, China)

Abstract: Based on some fundamental quantum chemical descriptors computed by PM3 hamiltonian. Fisher's linear discriminant functions in the development of classification models for the ability of polycyclic aromatic hydrocarbons (PAHs) to be directly cytotoxic to a cell line from the rainbow trout gill was developed using stepwise discriminant method. The functions obtained can be used for initial prediction of other PAHs. The prediction accuracy for the observed values is 100%. The main factors affecting cytotoxicity of PAHs are molecular weight (Mw), dipole moment (μ), energy of the highest occupied molecular orbital (E_{HOMO}), energy of the lowest unoccupied molecular orbital (E_{LUMO}) and ($E_{LUMO}-E_{HOMO}$)^2 of PAHs.

Key words: polycyclic aromatic hydrocarbons (PAHs), cytotoxicity, discriminant analysis, rainbow trout gill cell line

1 Introduction

Polycyclic aromatic hydrocarbons (PAHs), produced mainly by incomplete combustion of fossil fuels and natural materials (i. e. , forest fires, prairie fires, or agricultural burning), are ubiquitous pollutants with high toxicity (Blumenstock et al. , 2000; Niu, 2003; Teineman et al. , 2002). Many PAHs cause carcinogenicity, mutagenicity and genotoxicity (Jocob, 1996; Menzie et al. , 1992). Moreover, some studies show that quite a few PAHs cause cytotoxicity to animal cells in culture after they have been activated metabolically (Lee et al. , 1993; Schirmer et al. , 1998). The US-EPA has included 16 PAHs on the list of priority pollutants. Thus it is of great importance for the purpose of risk assessment of these widespread pollutants in the environment.

So far, for the PAHs occurred in the environment, their cytotoxicity data are lacking. On the other hand, because of large expenditures of money and time, it is difficult to determine the cytotoxicity for all PAHs. Thus it is necessary to model and predict their cytotoxicity using mathematical model. It is thus the purpose of this study to develop models by Fisher's discriminant analysis based on the determined cytotoxicity and to explain the molecular structural characteristics governing the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill based on the models. Despite of being almost 70 year old Fisher's discriminant analysis remains to be one of the most widely used methods of dimensionality reduction and classification (Sierra, 2002; Worth et al. , 2003). As quantum chemical descriptors can be easily obtained by computation, can clearly describe defined molecular properties, and are not restricted to closely related compounds, the development of models in which quantum chemical descriptors are used is of great importance (Chen et al. , 2001; Chen et al. , 2001).

2 Materials and methods

Schirmer et al. investigated the ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout (*Oncorhynchus mykiss*) gill, RTgill-W1 (Schirmer et al. , 1998). Three fluorescent indicator dyes were used to screen 16 PAHs for their ability to cause cytotoxic responses in RTgill-W1 cells. Alamar Blue (metabolic activity), 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA-AM; cell membrane integrity) and neutral red (lysosomal membrane integrity) were the indicator dyes and gave similar results. They found that 5 priority PAHs showed cytotoxic and no cytotoxicity was observed at the highest concentration tested for the others (Schirmer et al. , 1998).

The 16 PAHs with determined values constitute the training set of the study. Other 9 PAHs selected randomly from PAH congeners for which cytotoxicity was not determined constitute the prediction set.

Table 1   The PAHs under study and their cytotoxicity.

No.	Compounds	Cytotoxicity (Observed) [a] ($1^b/0^c$)	Cytotoxicity (Predicted) ($1^b/0^c$)	No.	Compounds	Cytotoxicity (Observed) [a] ($1^b/0^c$)	Cytotoxicity (Predicted) ($1^b/0^c$)
1	Naphthalene	1	1	14	Dibenz[a,h]anthracene	0	0
2	Acenaphthylene	1	1	15	Benzo[g,h,i]perylene	0	0
3	Acenaphthene	1	1	16	Indeno[1,2,3-cd]pyrene	0	0
4	Fluorene	1	1	17	9-Methylanthracene		0
5	Phenanthrene	1	1	18	9,10-Dimethylanthracene		0
6	Anthracene	0	0	19	9-Phenylanthracene		0
7	Fluoranthene	0	0	20	1-Methylphenanthrene		1
8	Pyrene	0	0	21	2-Methylphenanthrene		1
9	Benzo[a]anthracene	0	0	22	3-Methylphenanthrene		1
10	Chrysene	0	0	23	Cyclopenta[c,d]pyrene		0
11	Benzo[b]fluoranthene	0	0	24	Perylene		0
12	Benzo[k]fluoranthene	0	0	25	Naphthacene		0
13	Benzo[a]pyrene						

a. Observeds: Observed values determined by Schirmer et al. (1998); b. 1: Cytotoxicity was observed at the highest concentration tested; c. 0: No cytotoxicity was observed at the highest concentration tested.

MOPAC (2000) contained in the CS Chem3D Ultra (Ver. 6.0) was used to compute quantum chemical descriptors. The molecular structures were optimized using eigenvector following (Baker, 1986), a geometry optimization procedure within MOPAC 2000. The geometry optimization criteria GNORM was set at 0.1. A total of 11 MOPAC derived descriptors reflecting the overall character of the PAH molecules were used in this study. These are molecular weight (Mw), standard heat of formation (ΔHf), total energy (TE), electronic energy (EE), core-core repulsion energy (CCR), average molecular polarizability (α), dipole moment (μ), E_{HOMO}, E_{LUMO}, most positive net atomic charges on a hydrogen atom (q^{H+}), and largest negative atomic charge on a carbon atom (q^{C-}). The values of the selected molecular descriptors are listed in Table 2. The compound numbers in Table 2 correspond to those in Table 1. The unit of ΔHf is kilocalorie, and units of energy, charge, dipole and polarizability were electron volts (eV), atomic charge units (a.c.u) and atomic units (a.u.) respectively. In addition, three combinations of frontier molecular orbital energies, $E_{LUMO} - E_{HOMO}$, $(E_{LUMO} - E_{HOMO})^2$ and $E_{LUMO} + E_{HOMO}$ were also selected as predictor variables. The $E_{LUMO} - E_{HOMO}$ and $E_{LUMO} + E_{HOMO}$ can be related to absolute hardness and electronegativity respectively (Faucon et al., 1999). Furthermore, $E_{LUMO} - E_{HOMO}$, $E_{LUMO} + E_{HOMO}$, and $(E_{LUMO} -$ $E_{HOMO})^2$ were proven to be significant in toxicity and activity studies (Chen et al., 2001; Faucon et al., 1999).

3   Results and Discussion

When numerous variables are available for analysis, stepwise discriminant analysis can be used to identify the best subset of variables for classifying objects, and to build a discriminant model based on these variables (Sierra, 2002; Worth et al., 2003). In this present study, based on some fundamental quantum chemical descriptors computed by PM3 hamiltonian. The Fisher's linear discriminant functions in the development of classification models for the ability of polycyclic aromatic hydrocarbons (PAHs) to be directly cytotoxic to a cell line from the rainbow trout gill was developed using forward stepwise discriminant method. In forward stepwise analysis, all variables are evaluated in the first step to determine which one provides the most significant and unique discrimination between groups. Once this variable has been included in the model, all remaining variables are evaluated to determine which one provides the next best discrimination. The procedure continues until the addition of a new variable does not significantly improve the discrimination between groups. The Fisher's linear discriminant functions obtained from discriminant analysis are as follows:

$$C(Func0) = -1.216 \times 10^4 + 3.927 \times 10^{-1} Mw$$

$$-5.955 \times 10^1 \mu - 7.035 \times 10^3 E_{HOMO}$$
$$+ 7.290 \times 10^3 E_{LUMO} - 5.026 \times 10^2$$
$$(E_{LUMO} - E_{HOMO})^2 \qquad (1)$$

$$C(\text{Func1}) = -1.156 \times 10^4 + 3.141 \times 10^{-1} Mw$$
$$- 3.941 \times 10^1 \mu - 6.854 \times 10^3 E_{HOMO}$$
$$+ 7.10^1 \times 10^3 E_{LUMO} - 4.891 \times 10^2$$
$$(E_{LUMO} - E_{HOMO})^2 \qquad (2)$$

The discriminant functions may be used to make prediction for other structure similar PAHs. When C (Func0) > C (Func1), the result indicates that the predicted value is 0, i. e., there is no cytotoxicity for the PAHs at the highest concentration tested. Otherwise, the predicted value is 1 and cytotoxicity was observed for the PAHs at the highest concentration tested. Based on the discriminant functions, cytotoxicity for the other PAHs was predicted (Table 1). It can be seen from Table 1 that the prediction accuracy is 100%. It can be concluded from Equation(1) and Equation (2) that the main factors affecting cytotoxicity of PAHs are molecular weight (Mw), dipole moment (μ), the highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}) and ($E_{LUMO} - E_{HOMO}$)^2 of PAHs. The present study also shows that the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill increases with the increase of μ, E_{HOMO} and ($E_{LUMO} - E_{HOMO}$)^2. Increasing Mw and E_{LUMO} values of PAHs leads to decrease of the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill.

Table 2   Quantum chemical descriptors for PAHs in this study.

No. [*]	Mw	ΔHf	TE	EE	CCR	α	μ	E_{HOMO}	E_{LUMO}	q^{H+}	q^{C-}
1	128.173	40.674	-1307.300	-6639.940	5332.640	83.815	0.000	-8.835	-0.408	0.105	-0.101
2	152.200	73.129	-1543.170	-8572.720	7029.540	101.791	0.115	-9.055	-1.062	0.115	-0.121
3	154.210	38.844	-1575.330	-9067.530	7492.200	98.450	0.221	-8.589	-0.352	0.106	-0.094
4	166.222	48.868	-1693.533	-9871.993	8178.460	108.829	0.143	-8.842	-0.335	0.107	-0.106
5	178.230	55.026	-1811.900	-10911.100	9099.240	123.590	0.006	-8.740	-0.535	0.111	-0.102
6	178.233	61.662	-1811.624	-10791.631	8980.007	131.370	0.004	-8.248	-0.971	0.108	-0.100
7	202.260	80.012	-2048.100	-13135.000	11086.880	143.475	0.093	-8.725	-1.044	0.107	-0.111
8	202.260	64.141	-2048.790	-13281.800	11233.030	147.556	0.000	-8.249	-1.010	0.106	-0.102
9	228.290	74.463	-2316.290	-15643.600	13327.310	173.150	0.011	-8.328	-0.934	0.114	-0.102
10	228.290	70.854	-2316.440	-15747.100	13430.650	168.154	0.000	-8.496	-0.783	0.111	-0.102
11	252.320	93.630	-2552.740	-18266.640	15713.910	188.425	0.107	-8.663	-1.074	0.109	-0.105
12	252.320	178.311	-2552.640	-18086.360	15533.720	193.254	0.054	-8.400	-1.012	0.110	-0.107
13	252.315	92.090	-2552.807	-18366.034	15813.227	218.447	0.020	-7.672	-1.511	0.116	-0.104
14	278.350	87.926	-2820.930	-20959.100	18138.170	218.125	0.000	-8.377	-0.918	0.114	-0.102
15	276.337	84.381	-2790.422	-21395.948	18605.526	208.738	0.022	-8.122	-1.167	0.113	-0.099
16	276.340	104.929	-2789.540	-21023.690	18234.160	215.575	0.170	-8.235	-1.386	0.108	-0.108
17	192.260	54.639	-1961.230	-12458.800	10497.580	140.213	0.128	-8.134	-0.964	0.111	-0.101
18	206.287	48.088	-2110.820	-14209.400	12098.540	149.400	0.003	-8.023	-0.960	0.111	-0.101
19	254.331	89.446	-2583.580	-18882.800	16299.180	186.255	0.071	-8.199	-0.988	0.115	-0.104
20	192.260	45.519	-1961.580	-12455.600	10493.980	132.946	0.122	-8.645	-0.519	0.112	-0.111
21	192.260	45.622	-1961.620	-12348.400	10386.790	133.840	0.128	-8.719	-0.504	0.111	-0.103
22	192.260	53.461	-1961.610	-12382.700	10421.080	133.320	0.127	-8.635	-0.522	0.114	-0.105
23	226.280	148.575	-2282.410	-15697.780	13415.370	167.638	0.243	-8.297	-1.435	0.121	-0.111
24	252.320	82.025	-2553.240	-18604.700	16051.490	191.504	0.001	-7.987	-1.275	0.111	-0.100
25	228.293	84.335	-2315.860	-15449.900	13134.040	187.130	0.000	-7.871	-1.353	0.108	-0.100

[*] The compound numbers correspond to those in Table 1.

4   Conclusions

In this study, based on some fundamental quantum chemical descriptors computed by PM3 hamiltonian, Fisher's linear discriminant functions in the development of classification models for the ability of polycyclic aromatic hydrocarbons (PAHs) to be directly cytotoxic to a cell line from the rainbow trout gill was developed using stepwise discriminant method. The functions obtained can be used for initial prediction of other

PAHs and the prediction accuracy for the observed values is 100%. The main factors affecting cytotoxicity of PAHs are molecular weight (Mw), dipole moment (μ), the highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}) and ($E_{LUMO} - E_{HOMO}$)^2 of PAHs. The present study also shows that the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill increases with the increase of μ, E_{HOMO} and ($E_{LUMO} - E_{HOMO}$)^2. Increasing Mw and E_{LUMO} values of PAHs leads to decrease of the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill.

References

Blume nstock M, Zimmermann R, Schramm KW and Kettrup A. Infuence of combustion conditions on the PCDD/F-, PCB-, PCBz- and PAH-concentrations in the post-combustion chamber of a waste incineration pilot plant. Chemosphere, 2000, 40: 987 –993.

Niu JF, Chen JW, Martens D, Quan X, Yang FL, Kettrup A and Schramm KW. Photochemical behavior of polycyclic aromatic hydrocarbons absorbed to spruce (Picea abies (L.) Karst.) needles under sunlight irradiation. Environ Pollut, 2003, 123: 39 –45.

Teinemaa E, Kirso U, Strommen MR and Kamens RM. Atmospheric behaviour of oil-shale combustion fly ash in a chamber study. Atmos Environ, 2002, 36: 813 –824.

Jocob J. The significance of polycyclic aromatic hydrocarbons as environmental carcinogens. Pure Appl Chem, 1996, 68: 301 – 308.

Menzie CA, Potocki BB and Santodonato J. Exposure to the carcinogenic PAHs in the environment. Environ Sci Technol, 1992, 26: 1278 – 1284.

Lee LEJ, Clemons JH, Bechtel DG, Caldwell SJ, Han KB, Pasitschniak-Arts M, Mosser DD and Bols NC. Development and characterization of a rainbow trout liver cell line expressing cytochrome P450-dependent monooxygenase activity. Cell Biol Toxicol, 1993, 9: 279 –294.

Schirmer K, Dixon DG, Greenberg BM and Bols NC. Ability of 16 priority PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Toxicol, 1998, 127: 129 –141.

Sierra A. High-order Fisher's discriminant analysis. Pattern Recogn, 2002, 35: 1291 – 1302.

Worth AP and Croninb MTD. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. J Mol Struc-Theochem, 2003, 622: 97 –111.

Chen JW, Peijnenburg WJGM, Quan X, Chen S, Martens D, Schramm KW and Kettrup A. Is it possible to develop a QSPR model for direct photolysis half-lives of PAHs under irradiation of sunlight. Environ Pollut, 2001, 114: 137 – 143.

Chen JW, Quan X, Yan Y, Yang FL and Peijnenburg WJGM. Quantitative structure-property relationship studies on direct photolysis of selected polycyclic aromatic hydrocarbons in atmospheric aerosol. Chemosphere, 2001, 42: 263 – 270.

Baker J. An algorithm for the location of transition states. J Comput Chem, 1986, 7: 385 –395.

Faucon JC, Bureau R, Faisant J, Briens F and Rault S. Predicting of the fish acute toxicity from heterogeneous data coming from notification files. Chemosphere, 1999, 38: 3261 – 3276.

Niu JF and Yu G. The use of fisher's discriminant analysis in the development of classification models for the ability of PAHs to be directly cytotoxic to a cell line from the rainbow trout gill. Computers and Applied Chemistry, 2004, 21(6): 803 – 806.

多环芳烃(PAHs)对虹鳟(*Oncorhynchus mykiss*)鱼腮细胞系细胞毒性的 Fisher 判别分析

牛军峰,余刚

(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 算法计算得到的多环芳烃(PAHs)的量子化学参数,应用逐步判别分析法分析 PAHs 的细胞毒性,建立了能成功预测 PAHs 细胞毒性的 Fisher 线性判别函数,函数预测结果的正确识别率达到 100%。研究认为影响 PAHs 细胞毒性的主要因素是 PAHs 的分子量(Mw)、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})和($E_{LUMO}-E_{HOMO}$)^2。

关键词:多环芳烃(PAHs);细胞毒性;判别分析;虹鳟鱼腮细胞系

中图分类号:O

文献标识码:A

文章编号:1001-4160(2004)06-803-806