ELSEVIER

Journal of Molecular Structure: THEOCHEM 719 (2005) 191–200

THEO CHEM

www.elsevier.com/locate/theochem

QSAR model of the phototoxicity of polycyclic aromatic hydrocarbons

Fabiana Alves de Lima Ribeiro, Márcia Miguel Castro Ferreira*

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, 13081-970 Campinas, SP, Brazil

Received 17 August 2004; revised 26 January 2005; accepted 27 January 2005

Abstract

A Quantitative Structure–Activity Relationship (QSAR) study of 67 polycyclic aromatic hydrocarbons (PAHs) is performed and a prediction rule for the phototoxicity of these compounds is proposed. The geometry of all PAHs was optimized by semi-empirical method and used to calculate the following electronic descriptors: the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}) and the GAP ($E_{\mathrm{HOMO}}-E_{\mathrm{LUMO}}$) between them. The relationships between these molecular descriptors and the photo-induced toxicity are non-linear, and Gaussian type functions were used to linearize them. The data set was autoscaled and partial least squares (PLS) with leave-one-out cross-validation was used for building the regression model. After the model was validated, the phototoxicity was predicted for a set of molecules in which experimental measurements are unknown. Pentaphene, benzo[b]chrysene and dibenz[a.j]anthracene are among the compounds potentially phototoxic as predicted by the model. A new GAP range (7.2 ± 0.7 eV) is proposed for the classification of phototoxic compounds, and a larger cutoff suggested for the normalized lethal time as $\log(1/\mathrm{ALT})\leq-2.95$.

Keywords: QSAR; PLS; PAH; Phototoxicity; PAHs

1. Introduction

The increasing concern with the input of PAHs in the environment has stimulated the investigation of their properties and biological activities mainly because of carcinogenic and toxic potential. These chemicals are well known as water, sediment, soil and air contaminants [1–5]. They are predominant in urban and rural areas, due to the burning of wood and coal, exhaust of gasoline and diesel combustion engines, the smoking of tobacco, and other combustion processes. They may be produced from saturated hydrocarbons under oxygen-deficient conditions. Hydrocarbons with very low molecular masses, including even methane, may act as precursors for the polycyclic aromatic compounds. Low-molar-mass hydrocarbons form PAHs by pyrosynthesis, at temperatures above 500 °C at which C–C and C–H bonds are broken to form free radicals. These radicals undergo dehydrogenation and combine chemically to form a variety of aromatic ring structures [1,6,7].

Larger PAHs, containing at least four condensed rings are particularly relevant to the environment because most of them are carcinogenic, mutagenic and extremely resistant to enzymatic degradation [6–8]. Recent researches have demonstrated that the toxic effect of chemicals can be induced or enhanced by visible or UV radiation (phototoxicity) [9–12].

Phototoxicity has been of concern to the US Environmental Protection Agency (EPA) relative to the promulgation of adequately protective water quality criteria (WQC) [9,10]. All living organisms that live near to the Earth's surface are under the effect of the electromagnetic radiation from sunlight, and the phototoxicity is an important phenomenon to evaluate the environmental impact of some compounds.

Ultraviolet light of wavelengths 290–400 nm and visible radiation (400–780 nm), are sufficiently energetic to modify the chemical structure of the compounds, and even start some chemical reactions or to generate biologically active species [10,12]. A well-known example of these toxic effects caused by light exposure is the photodermatitis

---

* Corresponding author. Tel.: +55 19 3788 3102; fax: +55 19 3788 3023.

E-mail address: marcia@iqm.unicamp.br (M.M.C. Ferreira).

0166-1280/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2005.01.026

(a kind of skin eruption) produced by many plants such as lemon, lime and celery, after being exposed to UV light. These plants contain furocoumarins, phototoxic substances that after sunlight exposure cause hyperpigmentation and skin eruption [10,13]. Other compounds such as xanthenes and porphyrins are used as sunlight-activated insecticides due to their phototoxicity [14].

Upon sunlight exposure, PAHs undergo rapid structural modification generally via oxidation reactions. The modified products are in many cases more toxic than their parent compounds, as in case of anthracene [12]. The toxicity is due to the π-orbital system of PAHs, which strongly absorbs in the ultraviolet and visible regions of the solar spectrum [12,15].

Some studies about the phototoxicity of PAHs can be found in the literature, but its effect and mechanism are still unknown. Mallakin et al. [12] used anthracene to study a possible route for a simulated photoreaction. Anthracene absorbs strongly in the UV-A and UV-B spectral regions, and its central ring is more reactive what makes it susceptible for attack by molecular O_2. Further, anthracene photomodification pathway results in more than 20 subproducts, some of them very hazardous.

The phototoxicity of PAHs depends upon molecular properties that affect their bioaccumulation in the biological tissues, their ability to absorb the sunlight and the characteristics of their excited states. Of particular significance in environmental risk assessment are the triplet states of a chemical, which subsequently interact with molecular oxygen to create superoxide anion radicals and other reactive oxygen species that cause the cytotoxicity [16,17].

In aquatic environment, hydrophobic PAHs may accumulate in the biological tissues, and become toxic after exposure to UV and visible light [11]. Some studies about fluoranthene in aquatic organisms have shown that organisms under effect of 5 mg/L of fluoranthene exhibit some phototoxic behavior when exposed to very intense radiation (75.2 mW/cm^2), independent of the time of exposure. Phototoxicity depends directly of the dose and intensity of the absorbed radiation [9–11,15,17].

The data used in this work were taken from Newsted and Giesy [18], and are experimental values of phototoxicity of PAHs for *Daphnia magna*, expressed as the median adjusted lethal time, ALT (LT_{50} normalized to a constant concentration to eliminate, as much as possible, the chemical concentration as an external factor). These authors proposed a non-linear relationship between the phototoxicity and the energy of the triplet and singlet state of a set of PAHs obtained by spectroscopy. Mekenyan et al. [17] proposed a new approach for this relationship considering the phototoxicity as a cumulative result of internal factors as light absorbance and chemical stability, and external factors like exposure intensity and exposure energy. They utilized the GAP ($E_{\mathrm{HOMO}} - E_{\mathrm{LUMO}}$) as the only molecular descriptor to correlate with the potential phototoxicity, in a non-linear relationship. In this work, a new QSAR study is performed on the data from Newsted and Giesy [18] for predicting the phototoxicity of PAHs using besides GAP, the energies of HOMO and LUMO as molecular descriptors. Partial least squares (PLS) method was used for the regression analysis [19–21].

2. Methodology

2.1. Data set

In this work, 67 non-substituted PAHs containing 2–7 rings with 5 and 6 carbon atoms, were studied. The experimental data for phototoxicity [17] are listed in Table 1, together with the electronic descriptors calculated: E_{LUMO}, E_{HOMO} and GAP. The group of structures used to construct the model (training set) is shown in Fig. 1.

2.2. Descriptors calculations

The geometry of all molecules was optimized and the frontier orbital energies and the GAP expressed in eV were obtained from semi-empirical molecular orbital calculations using the method AM1 [22] implemented in Spartan software for UNIX [23].

2.3. Modeling and prediction

The QSAR model for phototoxicity was constructed with the PLS [19–21] method, on meancentered data and validated by leave-one-out cross-validation. The data analysis was performed by the programs Pirouette 2.02 (Infometrix) [24] and PLS Toolbox [25] for MATLAB (MathWorks Inc.) [26]. The goodness of the model was accessed by the following statistical parameters: the Prediction Error Sum of Squares, PRESS of validation, (Eq. (1)); the standard error of validation, SEV, (Eq. (2)); and finally by the standard and cross-validated correlation coefficients R^2 and Q^2, (Eqs. (3) and (4), respectively):

$$\mathrm{PRESS}_{\mathrm{val}} = \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{1}$$

$$\mathrm{SEV} = \sqrt{\frac{\mathrm{PRESS}}{n}} \tag{2}$$

$$R^2 = 1 - \left(\frac{\mathrm{PRESS}_{\mathrm{cal}}}{\sum_{i=1}^{n}((y_i - \hat{y}_i)^2)} \right) \tag{3}$$

$$Q^2 = 1 - \left(\frac{\mathrm{PRESS}_{\mathrm{val}}}{\sum_{i=1}^{n}((y_i - \hat{y}_i)^2)} \right) \tag{4}$$

where n is the number of compounds in the training set, y_i is the experimental value of the phototoxicity for the ith compound and $\hat{y}_i$ is the value predicted by the model built without compound i. $\mathrm{PRESS}_{\mathrm{cal}}$ is the PRESS calculated

Table 1
Molecular descriptors of the QSAR model, and the experimental and predicted values for phototoxicity of polycyclic aromatic hydrocarbons (PAHs)

	CAS name	E_{LUMO}	E_{HOMO}	GAP	Phototoxicity				Phototoxic Class		
					Exp.	Model 1	Model 2	Model 3	A	B	C
1	Anthracene	−0.8417	−8.1212	7.2795	−2.46		−2.53		1	1	+
2	Phenanthrene	−0.4085	−8.6171	8.2086	−3.20		−3.17		3	3	−
3	Naphthacene	−1.2321	−7.7488	6.5167	−2.89				1	2	+
4	Benz[a]anthracene	−0.8116	−8.2079	7.3963	−2.52		−2.61		1	1	+
5	Chrysene	−0.6762	−8.3697	7.6935	−2.54		−2.83		2	2	+
6	Triphenylene	−0.4532	−8.6584	8.2052	−3.20		−3.17		3	3	−
7	Pyrene	−0.9225	−8.0692	7.1467	−2.31		−2.50		1	1	+
8	Perylene	−1.1508	−7.8598	6.7090	−2.67				1	2	+
9	Benzo[a]pyrene	−1.1142	−7.9173	6.8031	−2.43				1	1	+
10	Benzo[e]pyrene	−0.8580	−8.2149	7.3569	−2.95	−2.42	−2.59		1	1	+
11	Dibenz[a. h]anthracene	−0.8041	−8.2570	7.4529	−2.44		−2.65		1	1	+
12	Benzo[ghi]perylene	−1.0662	−8.0235	6.9573	−2.36			−2.46	1	1	+
13	Fluoranthene	−0.9294	−8.6301	7.7007	−2.81		−2.81		1	2	+
14	Fluorene	−0.2088	−8.7109	8.5021	−3.20	−3.59	−3.38		3	3	−
15	Benzo[b]fluorene	−0.4880	−8.4783	7.9903	−3.16		−3.01		2	3	−
16	Benzo[k]fluoranthene	−1.1769	−8.3164	7.1395	−2.89	−2.44	−2.6		1	1	+
17	Benzo[a]fluorene	−0.5607	−8.3656	7.8049	−3.17		−2.92		2	3	−
18	Naphthalene	−0.2650	−8.7099	8.4449		−3.55	< −3.2	< −3.3			−
19	Benzo[c]phenanthrene	−0.6456	−8.4438	7.7982		−2.81		−2.58			+
20	Picene	−0.7209	−8.3487	7.6278		−2.62	−2.40	−2.50			+
21	Pentaphene	−0.8400	−8.2022	7.3622		−2.42		−2.39			+
22	Benzo[b]chrysene	−0.9948	−8.0511	7.0563		−2.41		−2.42			+
23	Dibenz[a. j]anthracene	−0.8736	−8.1916	7.3180		−2.41	−2.36	−2,42			+
24	Benzo[b]triphenylene	−0.8319	−8.2255	7.3936		−2.44	−2.40	−2.58			+
25	Benzo[c]chrysene	−0.6931	−8.3898	7.6967		−2.69		−2.54			+
26	Pentacene	−1.5500	−7.4414	5.8914		−3.57		< −3.3			−
27	Dibenzo[c. g]phenanthrene	−0.6732	−8.3498	7.6766		−2.67		−2.46			+
28	Benzo[a]naphthacene	−1.1857	−7.8407	6.6550		−2.72		2.68			+
29	Dibenzo[b. def]chrysene	−1.3630	−7.6784	6.3154		−3.12		−3.23			−
30	Dibenzo[def. mno]chrysene	−1.4067	−7.6315	6.2248		−3.22		< −3.3			−
31	Dibenzo[a.j]naphthacene	−1.1349	−7.9321	6.7972		−2.58		−2.55			+
32	Dibenzo[a.l]naphthacene	−1.1352	−7.9345	6.7993		−2.57		−2.54			+
33	Dibenzo[a.c]naphthacene	−1.1564	−7.9148	6.7584		−2.61		−2.58			+
34	Dibenzo[el]naphthacene	−0.8276	−8.2948	7.4672		−2.49		−2.42			+
35	Dibenzo[de.gr]naphthacene	−0.8336	−8.2774	7.4438		−2.47		−2.42			+
36	Dibenzo[g.p]chrysene	−0.8832	−8.1128	7.2296		−2.39		−2.39			+
37	Benzo[c]picene	−0.8279	−8.2597	7.4318		−2.46		−2.41			+
38	Dibenzo[b.k]chrysene	−1.1775	−7.8832	6.7057		−2.66		−2.65			+
39	Dibenzo[cl]chrysene	−0.7827	−8.2649	7.4822		−2.49		−2.40			+
40	Benzo[b]perylene	−1.1806	−7.8666	6.6860		−2.69		−2.64			+
41	Benzo[a]perylene	−1.4836	−7.5284	6.0448		−3.42		< −3.3			−
42	Dibenzo[de.mn]naphthacene	−1.5482	−7.4305	5.8823		−3.58		< −3.3			−
43	Naphtho[2.3-g]chrysene	−0.9944	−8.1177	7.1233		−2.40		−2.40			+
44	Benzo[h]pentaphene	−0.8089	−8.3009	7.4920		−2.51		−2.44			+
45	Benzo[a]pentacene	−1.4685	−7.5763	6.1078		−3.35		< −3.3			−
46	Coronene	−1.0021	−8.1438	7.1417		−2.39	−2.27	−2.39			+
47	Naphtho[1.2.3.4-def]chrysene	−1.0583	−8.0233	6.9650		−2.45		−2.45			+
48	Dibenzo[def.p]chrysene	−1.1022	−7.9553	6.8531		−2.53		−2.52			+
49	Benzo[rst]pentaphene	−1.1838	−7.8650	6.6812		−2.69		−2.65			+
50	Benzo[g]chrysene	−0.7660	−8.2705	7.5045		−2.51		−2.42			+
51	2.3:5.6-Dibenzophenanthrene	−0.9636	−8.0436	7.0800		−2.40		−2.41			+
52	Naphtho[2.1.8-qra]-naphthacene	−1.3169	−7.7299	6.4130		−3.00		−3.02			−
53	Dibenz[a.e]aceantrylene	−1.2806	−8.1423	6.8617		−2.55					+
54	Acenaphthylene	−0.9359	−8.9428	8.0069		−3.17					−

(*continued on next page*)

Table 1 (*continued*)

	CAS name	E_{LUMO}	E_{HOMO}	GAP	Phototoxicity				Phototoxic Class		
					Exp.	Model 1	Model 2	Model 3	A	B	C
55	Dibenzo[a.k]fluoranthene	−1.2992	−7.9434	6.6442		−2.74					+
56	Naphtho[2.3-k]fluoranthene	−0.9127	−7.9780	7.0653		−2.41					+
57	Dibenzo[k.mno]fluoranthene	−0.9755	−8.4000	7.4245		−2.49					+
58	1.2-Dihydroacenaphthylene	−0.2132	−8.4944	8.2812		−3.37					−
59	Benzo[c]fluorene	−0.6415	−8.2835	7.6420		−2.63					+
60	Benzo[ghi]fluoranthene	−0.9911	−8.6995	7.7084		−2.80					+
61	Benzo[a]aceanthrylene	−1.3219	−8.0850	6.7631		−2.63					+
62	Indeno[1.2.3.cd]pyrene	−1.2835	−8.1363	6.8528		−2.56					+
63	Indeno[1.2.3.cd]fluoranthene	−1.3350	−8.5435	7.2085		−2.54					+
64	Cyclopenta[cd]pyrene	−1.3123	−8.2727	6.9604		−2.52					+
65	Benzo[j]fluoranthene	−1.1767	−8.3165	7.1398		−2.44					+
66	Dibenz[e.k]acephenanthrylene	−1.0702	−8.2215	7.1513		−2.41					+
67	Benzo[b]fluoranthene	−0.9654	−8.6166	7.6512		−2.72					+
	SEV	0.1491									
	$\text{PRESS}_{\text{VAL}}$	0.3112									
	% Variance described	94.17									
	Q^2	0.7955									
	R^2	0.8411									
	Number of latent variables	1									

Experimental: from Ref. [17]. Model 1: predicted in this work. Model 2: predicted in Ref. [17]. Model 3: predicted by Ref. [29]. A: extracted from Ref. [18], where 1 means extremely toxic, 2 means moderately toxic and 3 means non-toxic. B: extracted from Ref. [17], where 1 means extremely toxic, 2 means moderately toxic and 3 means non-toxic. C: proposed by this work, where (+) means toxic and (−) means non-toxic.

when all compounds from the training set are included in the model.

3. Results and discussion

The frontier orbitals HOMO and LUMO can be used, respectively, as a measure of molecular capacity to donate or to accept an electron pair. The gap between them expresses the necessary energy to excite an electron from HOMO to LUMO, being proportional to the frequency of the absorbed photon ($\text{GAP} \propto \nu$) and to induce the phototoxicity in molecules. Compounds with larger GAP at constant light intensity exposure, absorb at lower wavelengths and may result in greater reactivity (exhibiting higher phototoxicity).

The absolute hardness (η) as well as the GAP, has an important influence on the photo-induced toxicity as pointed by Mekenyan et al. [17]. The hardness (η) is defined as the GAP/2 and it is known to be a measure of energy stabilization of an aromatic system and directly proportional to the PAHs stability [27,28].

Chemical stability and light absorption depend upon molecular structure, which can be well characterized by electronic descriptors. Small structural variations influence frontier orbital energies and are related with the energy required to induce the toxicity [17].

However, as seen previously [17,29] these relationships between molecular descriptors and photo-induced toxicity are not linear. Fig. 2(a)–(c) show the phototoxicity of PAHs increasing as a function of the molecular descriptors until reaches its maximum value, and then decreasing. A few exceptions are benzo[e]pyrene, fluoranthene, fluorene and benzo[k]fluoranthene which do not follow this general trend. The species with highest photo-induced toxicity (maximum in the curves) have their electronic descriptors values around 8.06 eV for E_{HOMO}, 0.91 eV for E_{LUMO} and 7.2 eV for GAP. These Gaussian type relationships between electronic descriptors and photo-induced toxicity result from the competing processes between the light absorbance, chemical stability, irradiation energy and irradiation intensity [17].

At small values of GAP, for PAHs in the left side of the curve (Fig. 2(c)), there are two factors that contribute to lower the toxicity. The first of them is the energy associated with the small GAP, which implies in reduction of the toxic effects and the second is the low hardness. Compounds with low values of hardness are more susceptible to structural modifications and are more easily degradable, resulting in compounds with less toxic effects.

Increasing the GAP should enhance the stability of PAHs resulting in greater phototoxicity, but external factors like the intensity of the radiation also affect the toxic potential [9,10,17] making these relationships non-linear. Under natural sunlight, the organisms are exposed to UV-B,

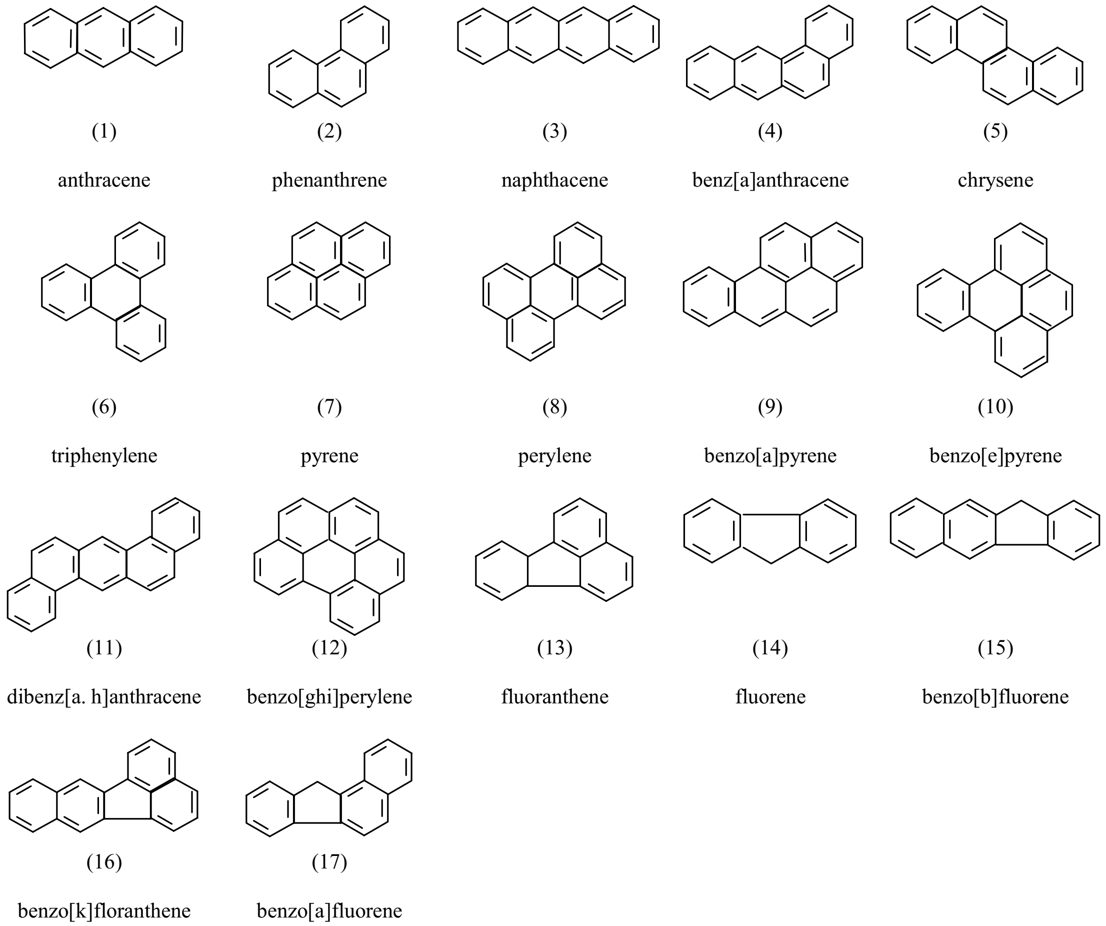
(1) anthracene
(2) phenanthrene
(3) naphthacene
(4) benz[a]anthracene
(5) chrysene
(6) triphenylene
(7) pyrene
(8) perylene
(9) benzo[a]pyrene
(10) benzo[e]pyrene
(11) dibenz[a. h]anthracene
(12) benzo[ghi]perylene
(13) fluoranthene
(14) fluorene
(15) benzo[b]fluorene
(16) benzo[k]floranthene
(17) benzo[a]fluorene

Fig. 1. Chemical structures of PAHs. These chemical structures were used to construct the PLS model, with exceptions of structures 10, 14 and 16, considered as outliers.

UV-A and visible radiation, with intensities ranging in the following order: UV-B < UV-A < visible, and molecules with smaller, intermediate and larger GAP absorb in the visible, UV-A and UV-B regions, respectively [17]. However, compounds with larger GAP absorb strongly at wavelengths that are absorbed in the outer atmosphere and do not reach the Earth's surface (only 0.1% of UV-B and 4.9% of UV-A reaches the earth), and so there is no enough intensity of the radiation to induce their phototoxicity [17].

The result of these competing effects may be observed in Fig. 2(a)–(c), where the compounds with smaller E_{HOMO}, higher E_{LUMO} and smaller GAP, which absorb at higher wavelengths and are more degradable, exhibiting small toxic effects. Some examples are naphthacene and perylene. In the central area of the curve are the compounds with intermediate E_{HOMO}, E_{LUMO} and GAP. They are less degradable and absorb at the spectral range present on the Earth's surface, favorable to the activation of their photo-induced toxicity. Some examples are anthracene, benz[a]anthracene, pyrene and benzo[ghi]perylene. Finally, those compounds with larger E_{HOMO}, and GAP and smaller E_{LUMO}, i.e. the PAHs with greater stability, absorb in the spectral range beyond that which reaches the Earth's surface, and therefore cannot absorb enough energy to become phototoxic. Some examples are: triphenylene and phenanthrene.

The Gaussian type functions proposed in previous work [29] were used to linearize these relationships (Eqs. (5)–(7)) and the results after linearization indicated that a good correlation especially with respect to the GAP could be obtained. After excluded the outliers benzo[e]pyrene, fluorene and benzo[k]fluoranthene from the data set, the correlation coefficient (R) between phototoxicity with E_{HOMO}, E_{LUMO} and GAP before linearization are, respectively, 0.59, 0.64 and 0.63. After linearization, the corresponding values of R for $E_{\mathrm{HOMO}}{}'$, $E_{\mathrm{LUMO}}{}'$ and GAP$'$ are 0.79, 0.88 and 0.91, respectively.

$$E'_{\mathrm{HOMO}} = \exp[-(E_{\mathrm{HOMO}} - 8.06)^2] \tag{5}$$

$$E'_{\mathrm{LUMO}} = \exp[-(E_{\mathrm{LUMO}} - 0.91)^2] \tag{6}$$

$$\mathrm{GAP}' = \exp[-(\mathrm{GAP} - 7.20)^2] \tag{7}$$

Initially, the experimental phototoxicity of 17 non-substituted PAHs (training set) were used to construct the regression model (Fig. 1 and Table 1), but three compounds presented high validation errors: benzo[e]pyrene, fluorene

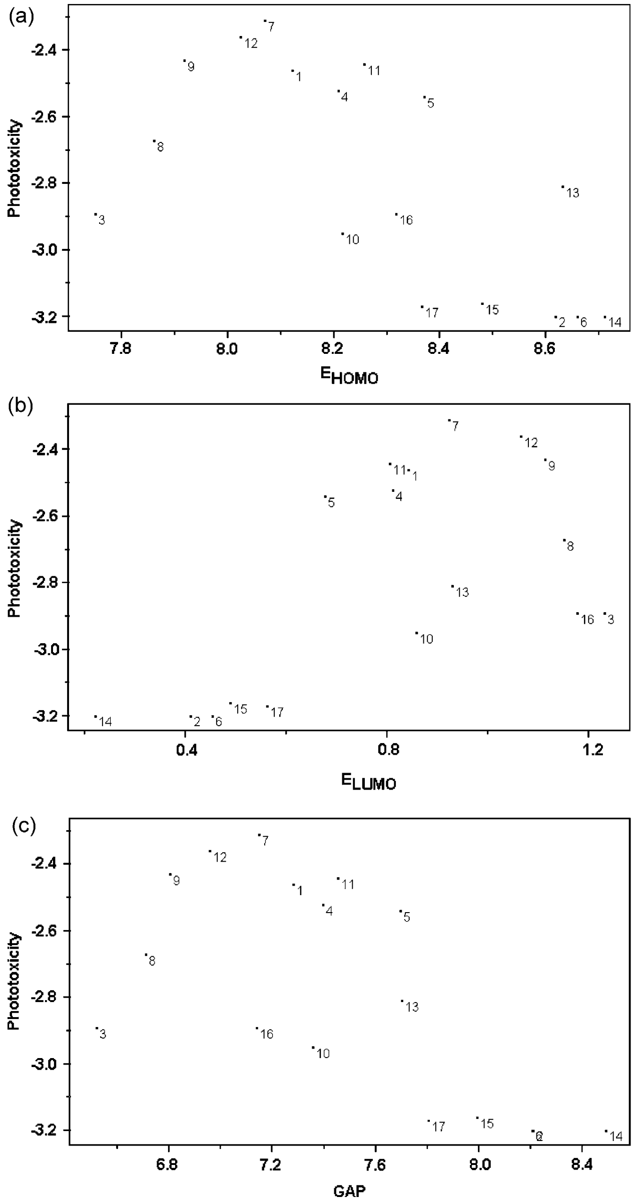

Fig. 2. Plots of phototoxicity versus E_{HOMO} (a), phototoxicity versus E_{LUMO} (b) and phototoxicity versus GAP (c), before linearization.

and benzo[k]fluoranthene. It was mentioned before that they do not follow the Gaussian type behavior in Fig. 2(a)–(c). These three compounds were considered as outliers and excluded from further analysis. The regression model was then constructed using the experimental phototoxicity of just 14 non-substituted PAHs. The regression analysis resulted in squared correlation coefficient $R^2=0.84$ and squared cross-validated correlation coefficient $Q^2=0.79$ with 1 latent variables describing 94.2% of total variance (Table 1). In a previous work, Ferreira [29] obtained $R^2=0.947$ with 2 latent variables for a data set consisting only of 11 PAHs, all of them with 6-membered rings. The model was built on autoscaled data. To obtain the usual regression equation, the regression coefficients were unscaled and the final results are presented in Eq. (8). The unscaling of the regression vector was performed as described by Ribeiro and Ferreira [30].

Table 2
Validation errors (%) for PLS model for phototoxicity

		Experimental phototoxicity	Predicted phototoxicity	Validation error (%)[a]
1	Anthracene	−2.46	−2.38	3.14
2	Phenanthrene	−3.20	−3.35	−4.69
3	Naphthacene	−2.89	−2.88	0.49
4	Benz[a]anthracene	−2.52	−2.43	3.67
5	Chrysene	−2.54	−2.70	−6.24
6	Triphenylene	−3.20	−3.35	−4.71
7	Pyrene	−2.31	−2.40	−4.02
8	Perylene	−2.67	−2.66	0.35
9	Benzo[a]pyrene	−2.43	−2.59	−6.38
11	Dibenz[a. h]anthracene	−2.44	−2.48	−1.65
12	Benzo[ghi]perylene	−2.36	−2.47	−4.75
13	Fluoranthene	−2.81	−2.75	2.01
15	Benzo[b]fluorene	−3.16	−3.01	4.66
17	Benzo[a]fluorene	−3.17	−2.78	12.39

[a] Validation errors were calculated from residuals (experimental—estimated value) obtained by leave-one-out crossvalidation, i.e. when the predicted compound is not included in the model.

$$\text{Phototoxicity} = -4.3355 + 0.4737\,E'_{\mathrm{HOMO}} + 0.3692\,E'_{\mathrm{LUMO}} + 1.1081\,\mathrm{GAP}' \qquad (8)$$

As shown in Table 2, the errors of validation obtained by leaving-one-out cross-validation, i.e. when the predicted compound is not included in the model, are all smaller than 6.4%, except for benzo[a]fluorene (12.4%). The plot of predicted versus experimental toxicity is in Fig. 3.

After validation, the constructed model was used to predict the phototoxicity of other 53 compounds (Fig. 4), whose descriptors are listed in the Table 1. The experimental values used to construct the model are found between −3.20 (phenanthrene and tryphenylene) and −2.36 (benzo[ghi]perylene). Some of the predicted values were

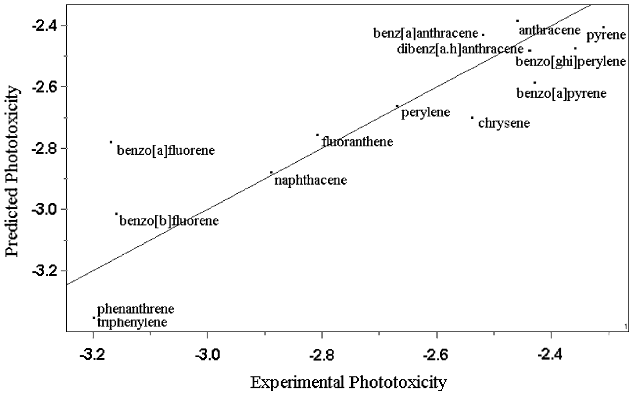

Fig. 3. Plot of experimental versus predicted values for phototoxicity for compounds from the training set.

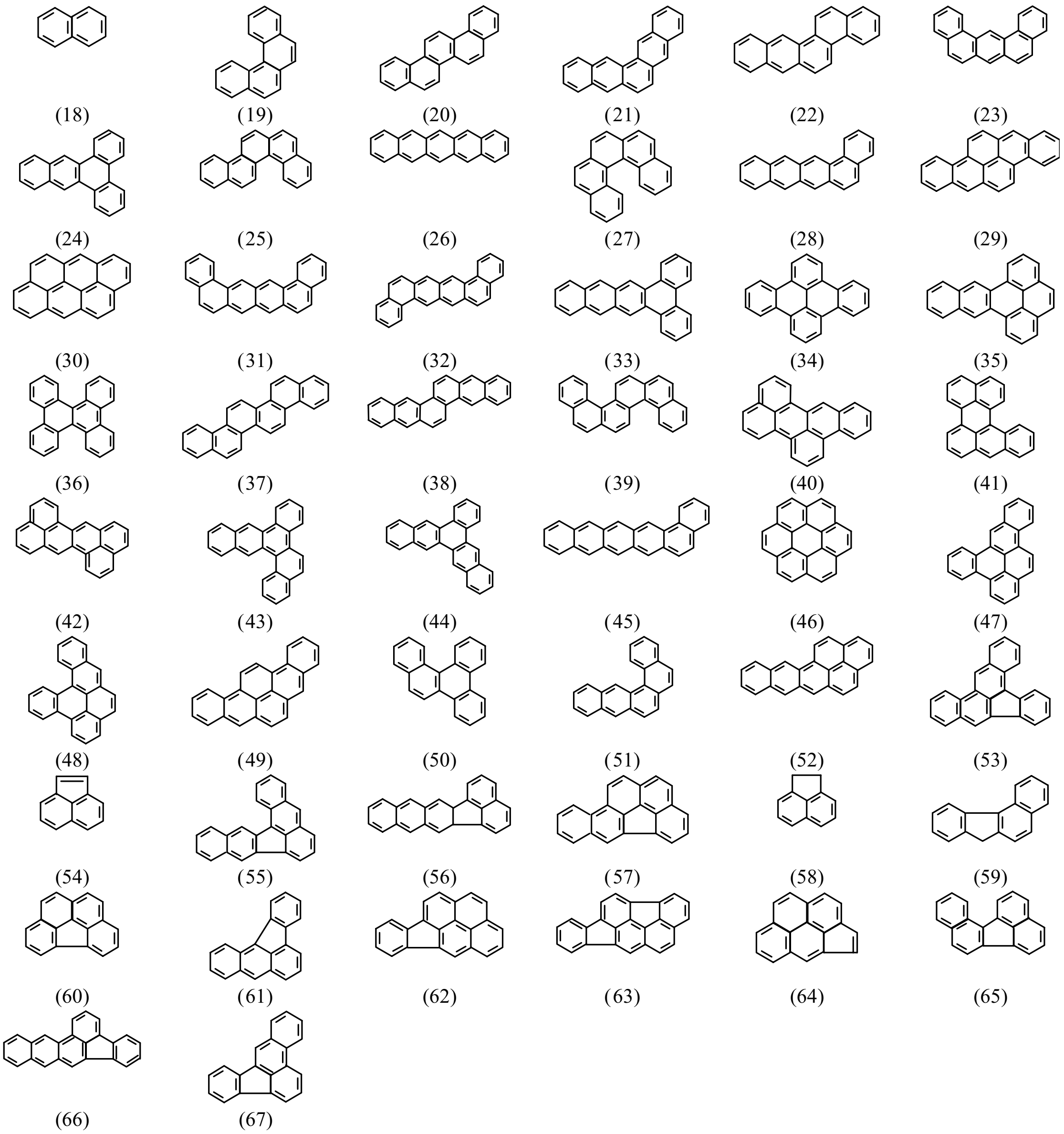
(18) (19) (20) (21) (22) (23)
(24) (25) (26) (27) (28) (29)
(30) (31) (32) (33) (34) (35)
(36) (37) (38) (39) (40) (41)
(42) (43) (44) (45) (46) (47)
(48) (49) (50) (51) (52) (53)
(54) (55) (56) (57) (58) (59)
(60) (61) (62) (63) (64) (65)
(66) (67)

Fig. 4. Chemical structures of PAHs whose values of phototoxicity were predicted by the PLS model.

out of this range but very close to the extreme values. In Table 1, Model 1 refers to the phototoxicity values predicted by the model proposed in this work. Models 2 and 3 refer, respectively, to the models proposed by Mekenian et al. [17] and Ferreira [29]. The predicted values are very similar for the three models. From the three compounds excluded as outliers, benzo[k]fluoranthene and benzo[e]pyrene were predicted as highly phototoxic.

In their experimental tests, Newsted and Giesy [18] observed three classes of phototoxic compounds, based on the time in seconds to induce the lethality to *D. magna* exposed to 120 μW/cm^2 UV-A and 25 μW/cm^2 UV-B light: class 1 (extremely toxic), 2 (moderately toxic) and 3 (non-toxic), described in Table 1. From their experimental observations [18], anthracene, naphthacene, benz[a]anthracene, pyrene, perylene, benzo[a]pyrene, benzo[e]pyrene, dibenz[a.h]anthracene, benzo[ghi]perylenefluoranthene, and benzo[k[fluoranthene were classified as extremely phototoxic compounds (class 1). Chrysene, benzo[b]fluorene and benzo[a]fluorene were classified as moderately phototoxic (class 2) and only phenanthrene, triphenylene, and fluorene were classified as non-toxic (class 3). Except chrysene (class 2), all the compounds in class 1 exhibit an adjusted lethal time (log(1/ALT)) ≥ -2.95. Then, under the laboratory conditions used by the authors [18], the threshold GAP of a PAH to induce phototoxicity is about 8.0 eV (as determined by both PM3 and AM1 methods). Phenanthrene, triphenylene and benzo[b]fluorene are the PAHs with a GAP$\geq$8.0 eV and do not exhibit any hazard behavior.

Mekenyan et al. [17] reexamined the experimental data from Newsted and Giesy [18], and proposed new external and internal factors to contribute for the toxicity of PAHs. Using a set of 16 PAHs, they proposed a range of GAP about 7.2 ± 0.4 eV for the most phototoxic molecules, that include anthracene, benzo[a]pyrene, dibenz[a.h]anthracene, pyrene, benz[a]anthracene, benzo[e]pyrene, benz[k]fluoranthene and benzo[ghi]perylene (Table 1). These molecules present adjusted lethal time (log(1/ALT)) of ≥ -2.52, with exceptions of benzo[k]fluoranthene (with -2.89) and benzo[e]pyrene (with -2.95).

Although being experimentally observed as extremely phototoxic by Newsted and Giesy [18], naphthacene, perylene and fluoranthene were classified by Mekenyan et al. [17] as moderately toxic (class 2). Besides the toxic behavior under laboratory conditions [17,18], naphthacene, on the extreme left in Fig. 2(b) and (c), is the largest PAH with linear character. As a consequence, has the lowest GAP value and stability, and has lower photo-induced toxicity under the environment sunlight. Perylene has been classified as extremely toxic under laboratory conditions, but exhibit a GAP value lower than the range proposed by Mekenyan et al. [17] for environment conditions, and so, has been classified as moderately toxic. Fluoranthene, also considered as in class 1 by Newsted and Giesy [18] present a larger GAP, characteristic of extremely stable compounds. However, at environment conditions cannot absorb enough energy to become phototoxic and has been considered by Mekenyan et al. as moderately toxic (class 2). Benzo[b]fluorene and benzo[a]fluorene, firstly considered as being from class 2 by Newsted and Giesy [18], were classified in class 3 by Mekenyan et al. [17] by the same reasoning as before.

From the 53 values of photo-induced toxicity estimated in the present work, five which are in common to Mekenyan's work present similar results (see Model 2 in Table 1). The model proposed by Mekenyan et al. [17] was based on the GAP energy, and was used to predict the phototoxicity of a group of just eight molecules. The model proposed here included two others variables, the energies of HOMO and LUMO, and was used to predict the phototoxicity of a large group of molecules. The estimated values of phototoxicity presented here are also in excellent agreement with the values predicted by Ferreira for 36 PAHs [29] with a similar model, as shown by Model 3 in Table 1, but with a data set consisting only 11 PAHs, all of them with just six-membered ring molecules. This work has also included into the data set some five-membered ring molecules, and the values of phototoxicity were estimated for a group of 50 molecules.

Considering the linearization described in Eqs. (5)–(7), this work introduces two new relationships to estimate phototoxic PAHs, based on the energies of their frontier orbitals. In Fig. 5(a)–(c) where the whole data set is taken into account (experimental and predicted), it is visible the existence of a gap between compounds with higher and smaller toxicity around the value of -2.95. These results lead us to suggest that molecules which present values of adjusted lethal time (log(1/ALT)) greater than -2.95, can be potentially phototoxic. As a consequence, the GAP range of 7.2 ± 0.4 eV suggested by Meckenyan et al. [17] must be increased to 7.2 ± 0.7 eV. A hierarchical cluster analysis carried out on the 17 structures from the training set, including the outliers molecules, using the variables $\exp[-(E_{\text{HOMO}}-8.06)^2]$, $\exp[-(E_{\text{LUMO}}-0.91)^2]$, and $\exp[-(\text{GAP}-7.20)^2]$, and the Euclidean distance as a similarity measurement (Fig. 6) shows two well-separated clusters corresponding to toxic and non-toxic compounds (benzo[a] fluorene and benzo[b]fluorene are exceptions). According to these results it is possible to classify the PAHs

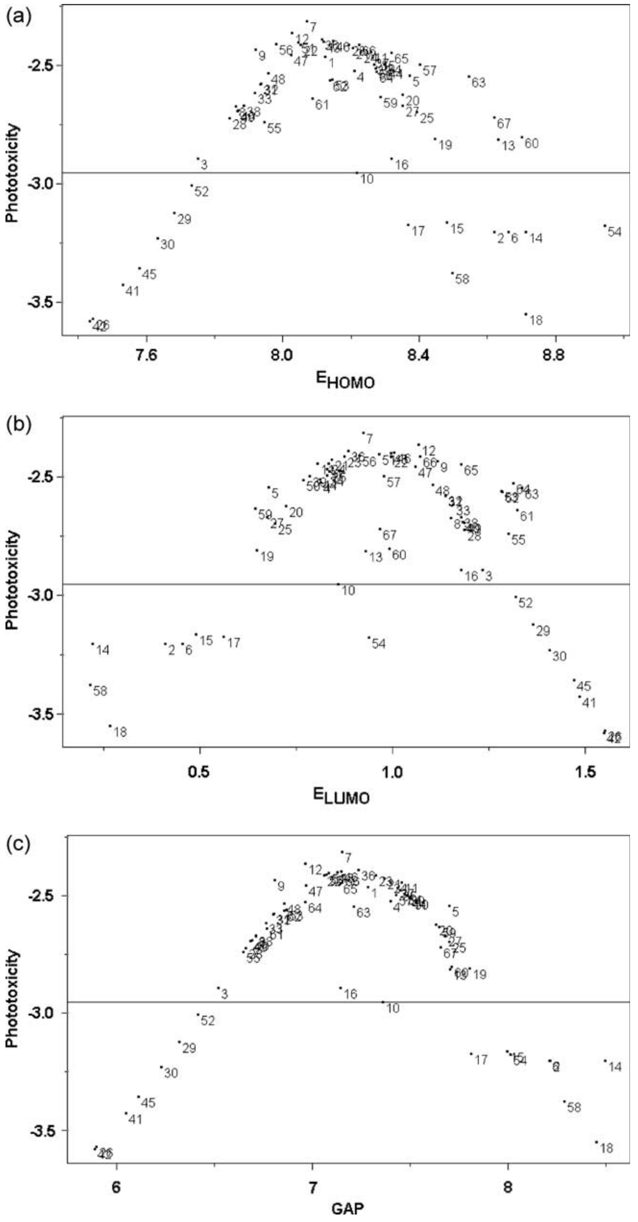

Fig. 5. Plots of phototoxicity versus E_{HOMO} (a), phototoxicity versus E_{LUMO} (b) and phototoxicity versus GAP (c) before the linearization, for all predicted and experimental values. The adjusted lethal time limit for photo-induced molecules is shown. The polycyclic aromatic hydrocarbons are identified by number (see Figs. 1 and 4).

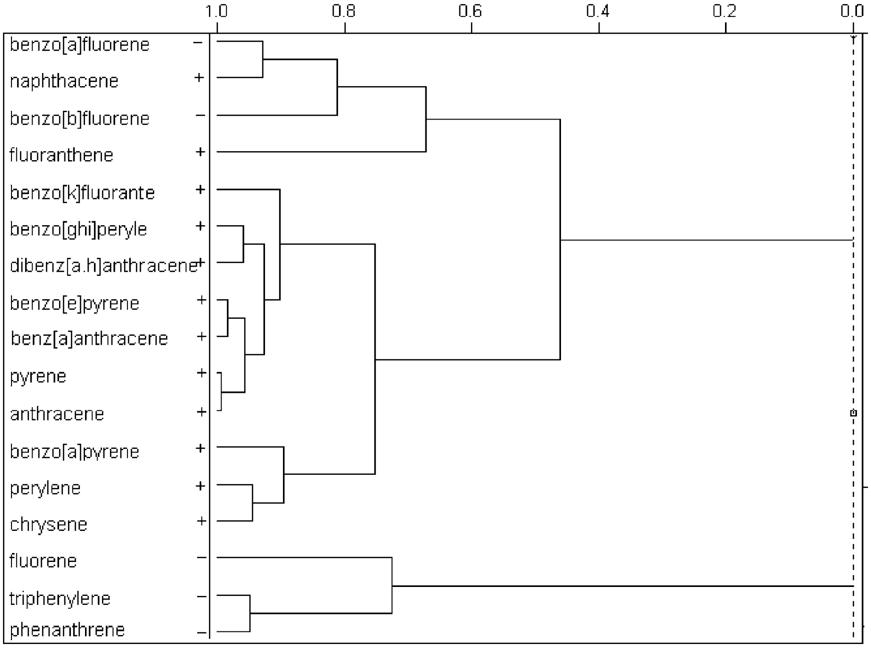

Fig. 6. Dendogram of the molecules samples from the training set, where (+) accounts for toxic and (−) for non-toxic molecules.

as photo/non-phototoxic and the results show that 41 more PAHs are expected to be phototoxic (see last column in Table 1). Hierarchical cluster on the whole data set (experimental and predicted by the model) was carried out and the results are in Fig. 7. Again, two well separated clusters appear corresponding to the two classes of molecules.

4. Conclusions

There are nowadays increasing interest in the QSAR studies, which consist of an important methodology that has been frequently used in medicinal chemistry, environmental science, material science etc. Frequently, the experimental determination of physical properties of substances is very

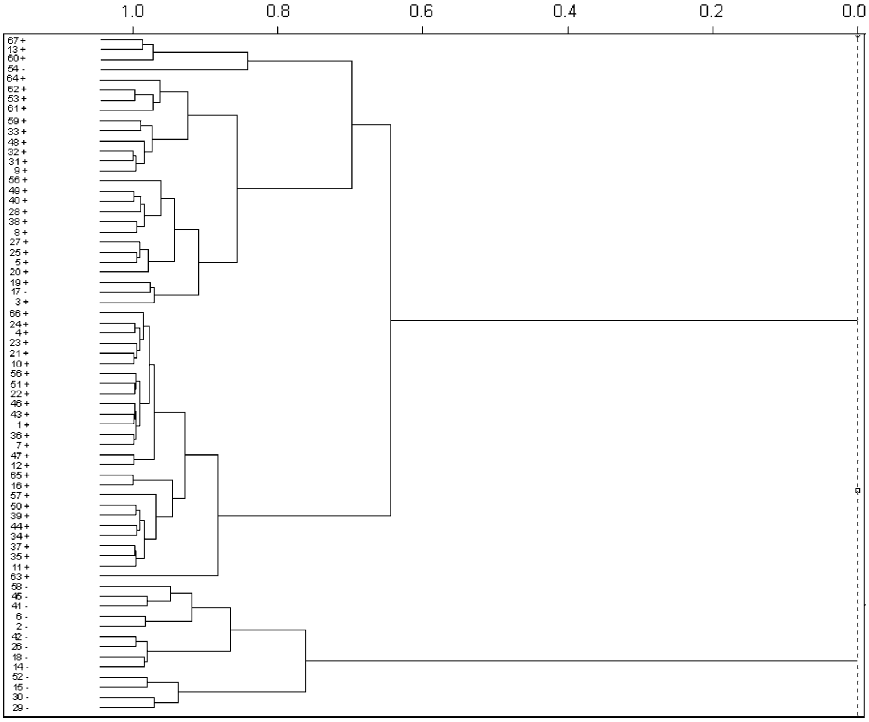

Fig. 7. Dendogram of the whole data set, where (+) accounts for toxic and (−) for non-toxic molecules. The polycyclic aromatic hydrocarbons are identified by number (see Figs. 1 and 4).

complex, time consuming and costly. Their handling may involve some risk, or may be not easy when compounds quickly degrade. On the other hand, the use of QSAR methods allows one to estimate through calculations and structural analysis which substances will be active or toxic, saving time money and effort.

In the presented work, the non-linear relationship between the electronic descriptors and toxicity was solved by using exponential transformations. The photo-induced toxicity for 53 PAHs was estimated from the QSAR study. Based on the results, a new scale for toxic compounds was proposed and the predicted values of phototoxicity allowed the classification of these molecules into toxic or non-toxic. It was also possible present the unscaled regression equation obtained by PLS model, by using a routine for unscaling the regression vector obtained from autoscaled data.

Acknowledgements

This work was supported by FAPESP, CAPES and FAEP. We thank Dr. Lucicleide R. Cirino for assistance with the theoretical calculations and valuable suggestions. Also CENAPAD, for computational support.

References

[1] A.D. Pereira Netto, J.C. Moreira, A.E.X.O. Dias, G. Arbilla, L.F.V. Ferreira, A.S. Oliveira, J. Barek, Quim. Nova 23 (2000) 765.
[2] D. Mackay, B. Hickie, Chemosphere 41 (2000) 681.
[3] P.D. Boehm, D.S. Page, E.S. Gilfillan, A.E. Bence, W. Burns, P.J. Mankiewicz, Environ. Sci. Technol. 32 (1998) 567.
[4] W. Burns, P.J. Mankiewicz, A.E. Bence, D.S. Page, K.R. Parkers, Environ. Toxicol. Chem. 16 (1997) 1119.
[5] I.J. Tolosa, J.M. Bayona, J. Albaiges, Environ. Sci. Technol. 30 (1996) 2495.
[6] S.E. Manahan, Environmental Chemistry, 6th ed., Lewis Publishers, Boca Ranton, 1994. chapter 7.
[7] C. Baird, Environmental Chemistry, 2nd ed., W. H. Freeman and Company, USA, 1995. chapter 14.
[8] A.F. Lehner, J. Horn, J.W. Flesher, J. Mol. Struct. 366 (1996) 203.
[9] G.T. Ankley, R.J. Erickson, G.L. Phipps, V.R. Mattson, P.A. Kosian, B.R. Sheedy, J.S. Cox, Environ. Sci. Technol. 29 (1995) 2828.
[10] D.J. Schaeffer, R.A. Larson, The Chemist July/August, 1999; 18.
[11] A.C. Hatch, G.A. Burton Jr., Environ. Pollut. 106 (1999) 157.
[12] A. Mallakin, D.G. Dixxon, B.M. Greenberg, Chemosphere 40 (2000) 1435.
[13] I.C. Weber, C.P. Davis, D.M. Greeson, J. Emerg. Med. 17 (1999) 235.
[14] T.B. Amor, G. Jori, Insect. Biochem. Mol. 30 (2000) 915.
[15] R.L. Peachey, D.G. Crosby, Mar. Environ. Res. 42 (1996) 359.
[16] G.D. Veith, O.G. Mekenyan, G.T. Ankley, D.J. Call, Environ. Sci. Technol. 29 (1995) 1267.
[17] O.G. Mekenyan, G.T. Ankley, G.D. Veith, D.J. Call, Chemosphere 28 (1994) 567.
[18] J.L. Newsted, J.P. Giesy, Environ. Toxicol. Chem. 6 (1987) 445.
[19] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1.
[20] P. Geladi, J. Chemometr. 2 (1988) 231.
[21] M.M.C. Ferreira, J. Braz. Chem. Soc. 13 (2002) 742.
[22] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[23] Spartan®, version 5.1.3, Wavefunction Inc., 1995–2001.
[24] Pirouette Multivariate Data analysis form IBM PC Systems, version 2.02, Infometrix: Seattle, WA, 1990–1996.
[25] PLS_Toolbox 2.0, Eigenvector Research: Manson, WA, 1998.
[26] MATLAB® for Windows, version 5.0, MathWorks Inc., 1995.
[27] Z. Zhou, R.G. Parr, J.F. Garts, Tetrahedron Lett. 29 (1988) 4843.
[28] Z. Zhou, R.G. Parr, J. Am. Chem. Soc. 111 (1989) 7371.
[29] M.M.C. Ferreira, Chemosphere 44 (2001) 125.
[30] F.A.L. Ribeiro, M.M.C. Ferreira, J. Mol. Struct. (Theochem) 663 (2003) 109.