J Mol Model (2012) 18:4121–4129
DOI 10.1007/s00894-012-1417-0

ORIGINAL PAPER

The use of quantum-chemical descriptors for predicting the photoinduced toxicity of PAHs

Jabir H. Al-Fahemi

Received: 13 January 2012 / Accepted: 26 March 2012 / Published online: 17 April 2012

Abstract The geometries of 19 polyaromatic hydrocarbons (PAHs) were fully optimized and calculated by a density functional method (B3LYP) with the 3-21G basis set. Various quantum chemical descriptors such as the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the difference in energy between those orbitals (E_{GAP}), electronegativity (χ), chemical potential (μ), chemical hardness (η), softness index (S), electrophilicity (ω), and polarizability (α) were employed along with physicochemical descriptors to construct useful quantitative structure–activity relationship (QSAR) models for the photoinduced toxicity of PAHs toward two aquatic species (*Daphnia magna* and *Scenedesmus vacuolatus*). E_{LUMO}, E_{HOMO}, E_{GAP}, S, χ, the molar refractivity (MR), and the molecular weight provide valuable information and play a significant role in the assessment of PAH phototoxicity. The resulting models are not expected to be useful per se for making genuine predictions for much larger test sets, but the various results do demonstrate the potential benefits of incorporating quantum-chemical descriptors into QSAR models for predicting the phototoxicity of PAHs.

Keywords Density functional method · QSAR · Multiple linear regression · Descriptors

Introduction

Polycyclic aromatic hydrocarbons (PAHs) have been found to exhibit toxic and hazardous properties [1]. These organic pollutants are known to become highly toxic in the presence of UV light [2–4]. UV radiation can enhance the toxicity of PAHs through photomodification to more toxic photoproducts [5, 6] and photosensitization reactions associated with the generation of reactive oxygen species [4, 7, 8]. Due to the continuous creation of large numbers of novel organic molecules nowadays, the development of models that can accurately and rapidly screen for the potential hazards of any hypothetical compounds in the aquatic environment is an important task. Quantitative structure–activity relationships (QSAR) are the most accurate predictive tools for initially evaluating the toxicities of chemical compounds using computer-aided models. There are a number of advantages to the use of QSAR for predicting toxicity. Key among these are the facts that toxicity can be predicted based on molecular structural parameters alone, and the models are easily automated [9, 10]. Such QSAR models are also significant to prior predictions. To obtain a good correlation it is important to use appropriate descriptors, whether they are theoretical, empirical, or derived from readily available experimental data on the structures. Many descriptors reflect simple molecular properties, so they can provide some meaningful insight into the physicochemical nature of the activity under investigation [11]. Quantum chemical descriptors, which can in principle express all of the electronic and geometric properties of molecules and their interactions, have been employed alone or in combination with conventional descriptors in many recent studies [11]. Indeed, some previous studies that used semiempirical quantum chemical methods found that the energy gap (E_{GAP}) between the highest occupied molecular orbital (E_{HOMO}) and the lowest unoccupied molecular orbital (E_{LUMO}) was an important factor in the photoinduced toxicity [3, 12–14]. In recent years, some comparative QSAR studies have shown that the use of parameters derived from density

J. H. Al-Fahemi (✉)
Department of Chemistry, University of Umm Al-Qura,
Makkah, Saudi Arabia
e-mail: jhfahemi@uqu.edu.sa

functional theory (DFT) instead of semiemperical methods can improve the accuracy of QSAR models [15–17]. In a recent paper, Wang et al. [18] applied descriptors calculated from DFT to the photoinduced toxicity of PAHs. They used E_{HOMO}, E_{LUMO}, E_{GAP}, the average molecular polarizability (α), the molecular molar volume (V), the vertical electron affinity in the singlet ground state (VEA_{S0}), the vertical ionization energy in the singlet ground state (VIE_{S0}), the lowest triplet excitation energy (E_{T1}), the vertical electron affinity in the lowest excited triplet state (VEA_{T1}), and the vertical ionization energy in the excited triplet state (VIE_{T1}) as descriptors. Using partial least squares regression, they established four- and six-descriptor models for the photoinduced toxicity of PAHs toward aquatic species. The former model yielded a correlation coefficient (R^2) of 0.82, whereas the second model was characterized by an R^2 value of 0.92.

The focus of the work described in the present paper was to investigate the utility of augmenting various quantum chemical quantities with physicochemical descriptors. Our main aim was to identify suitable molecular descriptors that might prove useful in predicting the photoinduced toxicity of PAHs.

Theory

Several important molecular properties such as electronegativity (χ), chemical potential (μ), chemical hardness (η), softness index (S), and electrophilicity (ω) are defined in density functional theory [19–21].

Electronegativity and chemical hardness are defined as

$$\chi = -\mu = -\left[\frac{\partial E}{\partial N}\right]_{v(r)} \tag{1}$$

and

$$\eta = \frac{1}{2}\left[\frac{\partial^2 E}{\partial N^2}\right]_{v(r)}, \tag{2}$$

respectively, where E and $v(r)$ are the electron energy and external potential of an N-electron system.

The operational definitions (χ and η) of these two quantities are

$$\chi = \frac{(I+A)}{2} \tag{3}$$

and

$$\eta = \frac{(I-A)}{2}, \tag{4}$$

respectively, where I and A are the ionization energy and the electron affinity of the chemical system (atom, ion, molecule), respectively.

Within the validity of Koopman's theorem, the energies of the frontier orbitals are given by

$$I \approx -E_{HOMO} \text{ and } A \approx -E_{LUMO}. \tag{5}$$

Equations 3 and 4 can be written as

$$\chi = 0.5(E_{HOMO} - E_{LUMO}) \tag{6}$$

and

$$\eta = 0.5(E_{LUMO} - E_{HOMO}). \tag{7}$$

The softness of a chemical system is a measure of its chemical reactivity. It is the reciprocal of the chemical hardness of the system:

$$S = \frac{1}{\eta}. \tag{8}$$

It is also proposed that the electrophilicity (ω) is a function of the square of the electronegativity (chemical potential) divided by the hardness [21]:

$$\omega = \frac{\mu^2}{2\eta}. \tag{9}$$

This study also included E_{HOMO}, E_{LUMO}, E_{GAP}, hydration energy (E_H), molecular weight (M), octanol water partition coefficients (log$P$), the surface area grid (σ), molecular volume (V), molar refractivity (MR), and molecular polarizability (α) as descriptors.

Finally, a shape descriptor (X) was also tested as a descriptor. X was found to be a promising descriptor in our previous work [22] that aimed to predict octanol–water partition coefficients. X was defined as

$$X = 36\pi\left(\frac{V^2}{\sigma^3}\right). \tag{10}$$

Values of X range from unity (for a sphere) down to zero.

Materials and methods

Data set

The data on the phototoxicities of 14 PAHs toward *Daphnia magna* and 12 PAHs toward *Scenedesmus vacuolatus* that were used in this study were those reported by Wang et al. [18]. The structures and the sequence numbers of the PAHs of interest in this work are shown in Fig. 1.

Quantum chemical calculations

All calculations performed in this investigation were done using the HyperChem program. The initial structures of all of the PAHs were constructed using the software CS Chem3D

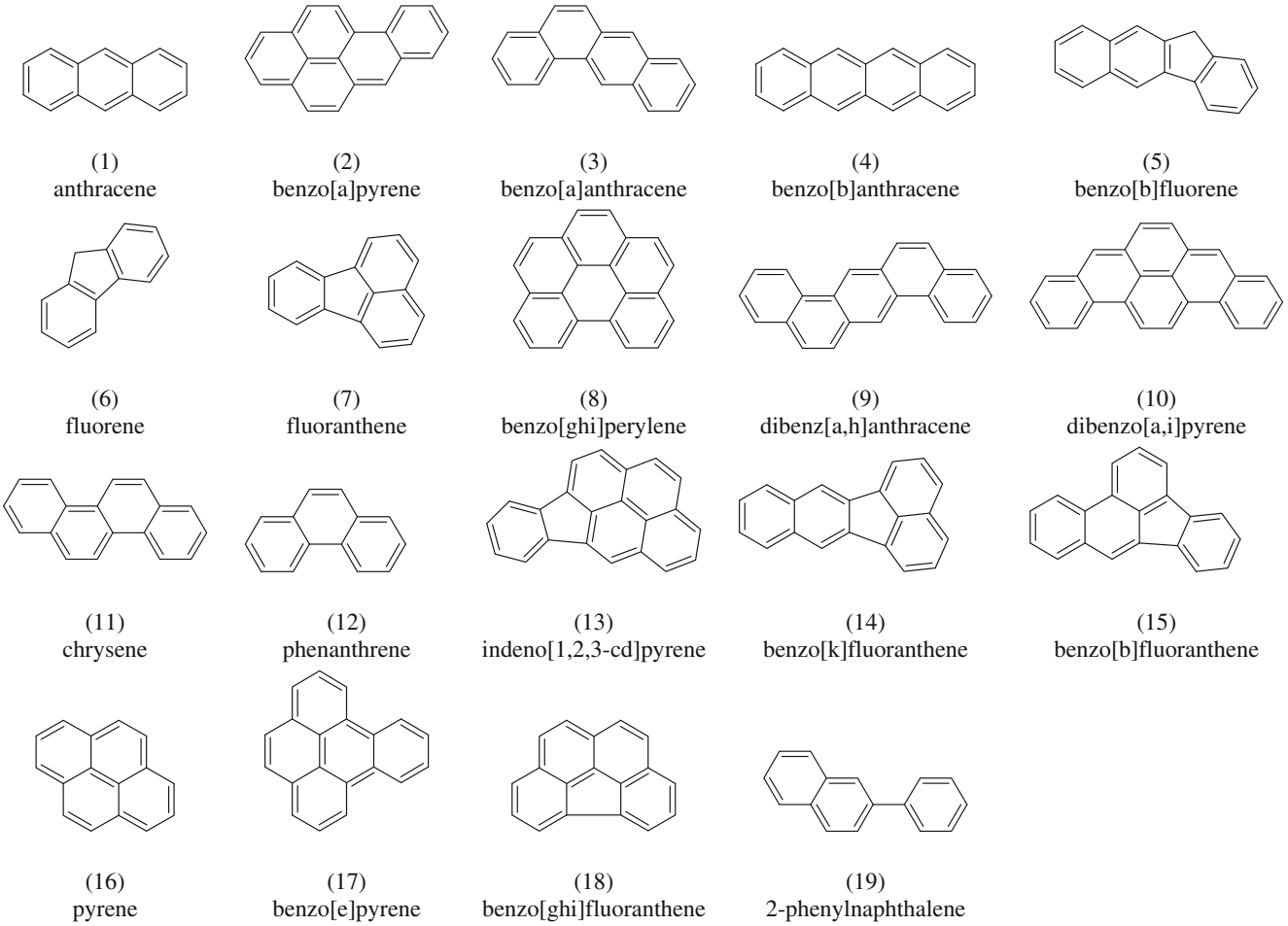

Fig. 1 Chemical structures of the molecules studied in this work

(Ultra 10.0, Chemoffice 2006, Cambridgesoft Corporation, Cambridge, MA, USA). To save computational time, the initial geometries of PAHs were preoptimized using the semi-empirical PM3 Hamiltonian. The resulting geometries were then reoptimized using DFT methods with the 3-21G basis set in combination with the potential B3-LYP.

Descriptor selection

The photoinduced toxicity of PAHs depends on molecular properties that affect their bioaccumulation in biological tissues, their ability to absorb light, and the characteristics of their excited states. The frontier orbitals—the HOMO and the LUMO—can be employed, respectively, as a measure of the molecule's capacity to donate or accept an electron pair. The gap between them is the energy required to excite an electron from the HOMO to the LUMO (which is proportional to the frequency of the absorbed photon) and to induce phototoxicity. Molecules with larger E_{GAP} values absorb at lower wavelengths and may exhibit higher phototoxicities under constant exposure to light [23]. In this study, the values of E_{HOMO}, E_{LUMO}, and E_{GAP} were obtained directly from DFT computations. However, the values for the electronegativity (χ), chemical potential (μ), chemical hardness (η), softness index (S), and electrophilicity (ω) were calculated using the equations given above and other values were obtained from DFT calculations. The calculated values of the molecular descriptors used in this study are presented in Table 1.

Statistical analysis

Multiple linear regression (MLR) was employed to generate linear models, and it was carried out with SPSS [24]. Because of the large number of molecular descriptors considered, principal component analysis (PCA) was performed within XLSTAT to determine the useful descriptors. After eliminating unsuitable descriptors, a stepwise multiple linear regression procedure based on the forward-selection and backward-elimination methods was employed to determine the best regression models. Statistical outliers in this study were considered to be molecules with absolute standardized residuals of >2. The quality of each MLR model was assessed based on the coefficient of determination (R^2), the adjusted R^2 (R_{adj}^2), the standard error of the estimate (s), Fisher's criterion (F), and the

Table 1 Calculated descriptors for the studied molecules

Molecule	E_H	E_{HOMO}	E_{LUMO}	E_{GAP}	M	log$P$	σ	V	MR	α	X	η	S	μ	χ	ω
Anthracene	−2.70	−5.25	−1.56	3.69	178.23	1.75	365.12	584.58	67.35	22.80	0.79	1.84	0.54	−3.41	3.41	10.72
Benzo[*a*]anthracene	−3.03	−5.36	−1.49	3.88	228.29	1.82	422.91	704.73	85.55	28.98	0.74	1.94	0.52	−3.43	3.43	11.38
Benzo[*a*]pyrene	−2.83	−5.15	−1.71	3.44	252.32	1.36	431.48	739.22	93.43	31.68	0.77	1.72	0.58	−3.43	3.43	10.10
Benzo[*b*]anthracene	−3.08	−4.88	−2.02	2.86	228.29	1.82	426.61	711.99	85.55	28.98	0.74	1.43	0.70	−3.45	3.45	8.53
Benzo[*b*]fluorene	−2.30	−5.57	−1.11	4.46	216.28	1.69	420.28	686.39	79.83	27.34	0.72	2.23	0.45	−3.34	3.34	12.44
Benzo[*e*]pyrene	−2.80	−5.47	−1.36	4.11	252.32	1.36	424.34	731.46	93.43	31.68	0.79	2.05	0.49	−3.41	3.41	11.97
Benzo[*ghi*]perylene	−2.70	−5.27	−1.63	3.63	276.34	0.91	441.2	772.11	101.30	34.39	0.79	1.82	0.55	−3.45	3.45	10.80
Chrysene	−2.96	−5.56	−1.22	4.35	228.29	1.82	417.59	698.52	85.55	28.98	0.76	2.17	0.46	−3.39	3.39	12.48
Dibenz[*a*,*h*]anthracene	−3.38	−5.43	−1.43	4.00	278.35	1.90	481.78	825.30	103.74	35.16	0.69	2.00	0.50	−3.43	3.43	11.78
Dibenzo[*a*,*i*]pyrene	−3.26	−5.08	−1.78	3.30	302.38	1.44	489.61	857.76	111.62	37.87	0.71	1.65	0.61	−3.43	3.43	9.73
Fluoranthene	−2.54	−5.83	−1.70	4.13	202.26	1.29	385.63	631.32	75.23	25.50	0.79	2.06	0.48	−3.77	3.77	14.65
Fluorene	−1.86	−5.80	−0.62	5.18	166.22	1.61	355.89	559.19	61.63	21.15	0.78	2.59	0.39	−3.21	3.21	13.38
Phenanthrene	−2.61	−5.78	−0.93	4.85	178.23	1.75	359.06	578.51	67.35	22.80	0.82	2.42	0.41	−3.35	3.35	13.62
Pyrene	−2.41	−5.36	−1.44	3.92	202.26	1.29	375.38	619.30	75.23	25.50	0.82	1.96	0.51	−3.40	3.40	11.36
Benzo[*b*]fluoranthene	−2.97	−5.80	−1.68	4.12	252.32	1.36	440.31	751.93	93.43	31.68	0.75	2.06	0.49	−3.74	3.74	14.37
Benzo[*k*]fluoranthene	−2.99	−5.45	−1.66	3.79	252.32	1.36	447.39	756.18	93.43	31.68	0.72	1.89	0.53	−3.55	3.55	11.96
Indeno[1,2,3-cd]pyrene	−2.85	−5.40	−1.98	3.42	276.34	0.91	455.62	789.24	101.30	34.39	0.74	1.71	0.58	−3.69	3.69	11.67
Benzo[*ghi*]fluoranthene	−2.40	−5.96	−1.79	4.18	226.28	0.83	401.94	670.28	83.11	28.21	0.78	2.09	0.48	−3.87	3.87	15.68
2-Phenylnaphthalene	−3.20	−5.65	−1.18	4.47	204.27	2.28	409.97	665.60	77.67	26.28	0.73	2.24	0.45	−3.41	3.41	13.03

Table 2 Correlation coefficient matrix for phototoxicity toward *Daphnia magna* and the calculated descriptors

	logEC_{50}	E_H	E_{HOMO}	E_{LUMO}	E_{GAP}	M	log$P$	σ	V	MR	α	X	η	S	μ	χ	ω
logEC_{50}	1																
E_H	0.696	1															
E_{HOMO}	−0.594	−0.588	1														
E_{LUMO}	0.744	0.640	−0.769	1													
E_{GAP}	0.721	0.656	−0.923	0.956	1												
M	−0.842	−0.726	0.547	−0.542	−0.578	1											
log$P$	0.344	−0.256	−0.049	0.242	0.168	−0.277	1										
σ	−0.775	−0.769	0.539	−0.507	−0.553	0.961	−0.041	1									
V	−0.814	−0.766	0.555	−0.533	−0.576	0.987	−0.139	0.992	1								
MR	−0.844	−0.753	0.556	−0.551	−0.587	0.999	−0.246	0.965	0.990	1							
α	−0.840	−0.743	0.551	−0.542	−0.580	1.000	−0.249	0.966	0.991	1.000	1						
X	0.392	0.577	−0.337	0.255	0.308	−0.605	−0.421	−0.799	−0.719	−0.614	−0.620	1					
η	0.721	0.656	−0.923	0.956	1.000	−0.578	0.168	−0.553	−0.576	−0.587	−0.580	0.308	1				
S	−0.605	−0.605	0.934	−0.915	−0.980	0.513	−0.109	0.505	0.521	0.523	0.516	−0.311	−0.980	1			
μ	0.453	0.298	−0.007	0.645	0.391	−0.194	0.319	−0.149	−0.170	−0.197	−0.189	−0.005	0.391	−0.314	1		
χ	−0.453	−0.298	0.007	−0.645	−0.391	0.194	−0.319	0.149	0.170	0.197	0.189	0.005	−0.391	0.314	−1.000	1	
ω	0.517	0.535	−0.991	0.677	0.863	−0.511	0.010	−0.509	−0.522	−0.519	−0.516	0.334	0.863	−0.889	−0.126	0.126	1

significance of Fisher's F statistic (p). The models were validated internally by leave-one-out (LOO) cross-validation. In this method, one molecule was eliminated from the data set at random in each cycle, and the model was then built using the other molecules. The model thus formed was used to predict the phototoxicity of the eliminated molecule. This process was repeated until all of the compounds had been eliminated once. The cross-validated R_{cv}^2 for the model was determined based on the predictive ability of the model; the higher the value of R_{CV}^2 (>0.5), the better the predictivity of the model.

Results and discussion

Set A (phototoxicity of 14 PAHs toward *Daphnia magna*)

The correlation coefficients (R) between the phototoxicity toward *Daphnia magna* and the calculated descriptors are presented in Table 2. High values of the correlation coefficient are important, regardless of whether they are positive or negative values. PCA results in a mathematical transformation of the original data matrix. It was used here as a visualization method for investigating the individual correlation between logEC_{50} and the molecular descriptors. In general, the data presented in Table 2 were abstracted into two factors, F1 and F2, as shown in Fig. 2. The correlation circle shows a projection of the initial parameters (descriptors) into factor space. When two parameters are close to each other, they are strongly positively correlated. On the other hand, if the parameters are on opposite sides of the center point, they are significantly negatively correlated. Orthogonal parameters are not correlated.

Figure 2 shows that the descriptors molecular weight, surface area grid, molecular volume, molar refractivity, and polarizability are very close to each other but located on the opposite side to logEC_{50}. These descriptors are therefore significantly negatively correlated with logEC_{50} ($R \approx -0.8$). On the other hand, hydration energy, E_{LUMO}, E_{GAP}, and chemical hardness are relatively close to logEC_{50} and show positive correlations with it ($R \approx 0.7$). E_{HOMO}, softness index, chemical potential, electronegativity, and electrophilicity show acceptable correlations with logEC_{50} ($0.5 \leq |R| \leq 0.6$). log$P$ and the molecular shape descriptor (X) exert only minor effects on logEC_{50} considering their low correlation coefficients, as shown in Table 2.

After eliminating inappropriate descriptors (log$P$ and X), we generated various linear models using the experimental values for photoinduced toxicity (logEC_{50}) as the dependent variable and the molecular structure descriptors as the independent variables. We only report the best equations below.

Using only two descriptors, we found that the best model involved only the molar refractivity and the energy of the lowest unoccupied molecular orbital:

$$\begin{aligned} \log EC_{50} &= 7.690 - 0.0528MR + 1.368E_{LUMO} \\ n &= 14, R^2 = 0.824, R^2_{adj} = 0.792, s = 0.571, \\ F &= 25.781, p_F = 7.1 \times 10^{-5}. \end{aligned} \quad (11)$$

The t values[1] for MR and E_{LUMO} were −4.118 and 2.640, indicating that both descriptors were important, although MR was the most important. The negative value of t

[1] For a molecular descriptor to be suitable for use as a predictor, its t value should satisfy $|t| > 2$.

indicated that molar refractivity was negatively correlated with logEC_{50}. On the other hand, E_{LUMO} was positively correlated with logEC_{50}. These results were consistent with the PCA results. The observed values of photoinduced toxicity toward *Daphnia magna* are compared numerically in Table 3 to the predictions obtained using our preferred linear model (11). The correlation between these predictions and the observed values of phototoxicity toward *Daphnia magna* is shown graphically in Fig. 3.

As judged from the correlation coefficients, the statistical quality of model (11) is similar to that of the model reported by Wang et al. [18] who obtained an R^2 value of 0.820 using four descriptors. However, model (11) is more statistically significant than the model reported in [18] due to its low value for the significance of Fisher's F statistic.

We used LOO cross-validation as an internal test of the quality of the two-descriptor model. The model's performance was good and was characterized by an $R_{CV}{}^2$ value of 0.669. In this model, dibenzo[*a*,*i*]pyrene had a standardized residual of 2.099. Excluding that molecule led to a better value for the cross-validation coefficient ($R_{CV}{}^2$= 0.772), but there is of course no justification for this reselection of the data set a posteriori.

The other model that was found to be useful was based on five descriptors and took the form

$$\begin{aligned}\log EC_{50} &= -2.057 - 0.366E_H - 3.614E_{HOMO} - 3.592M \\ &\quad + 3.253S - 4.435\chi \\ n &= 14, R^2 = 0.929, R^2{}_{adj} = 0.884, s = 0.426, \\ F &= 20.907, p_F = 2.1 \times 10^{-4}.\end{aligned} \quad (12)$$

Table 3 Observed and predicted values of phototoxicity toward *Daphnia magna* for the set of 14 PAHs. Predictions were made using model (11)

Molecule	logEC_{50} (observed)	logEC_{50} (predicted)
Anthracene	1.80	1.99
Benzo[*a*]anthracene	0.62	1.14
Benzo[*a*]pyrene	0.59	0.42
Benzo[*b*]anthracene	1.01	0.41
Benzo[*b*]fluorene	1.61	1.95
Benzo[*e*]pyrene	0.11	0.90
Benzo[*ghi*]perylene	−0.32	0.11
Chrysene	1.24	1.51
Dibenz[*a*,*h*]anthracene	0.30	0.26
Dibenzo[*a*,*i*]pyrene	0.36	−0.64
Fluoranthene	1.29	1.39
Fluorene	4.23	3.58
Phenanthrene	3.43	2.86
Pyrene	1.36	1.75

However, the internal predictive power of this five-descriptor model was very poor ($R_{CV}{}^2$=0.376). In this model, benzo(*b*)anthracene and fluoranthene had standardized residuals of −2.040 and 2.850, respectively. Excluding those molecules improved the model's predictivity ($R_{CV}{}^2$=0.86).

The values of t for E_H, E_{HOMO}, M, S, and χ were −0.366, −3.614, −3.592, 3.253, and −4.435, respectively. These values indicated that all of these descriptors are important except for the hydration energy. Excluding E_H from the regression model improved the statistical quality of the

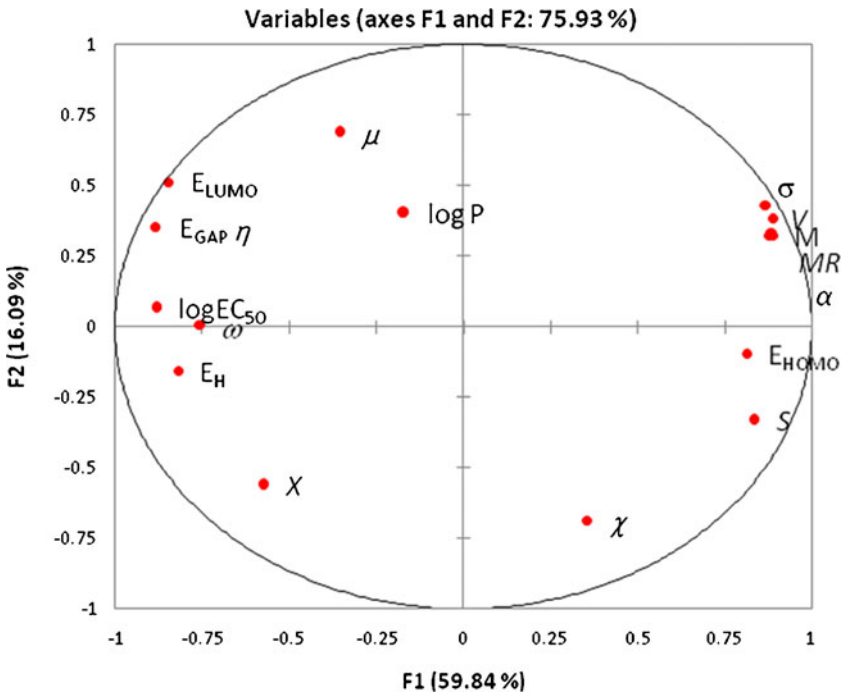

Fig. 2 Correlation circle for phototoxicity toward *Daphnia magna* and calculated descriptors

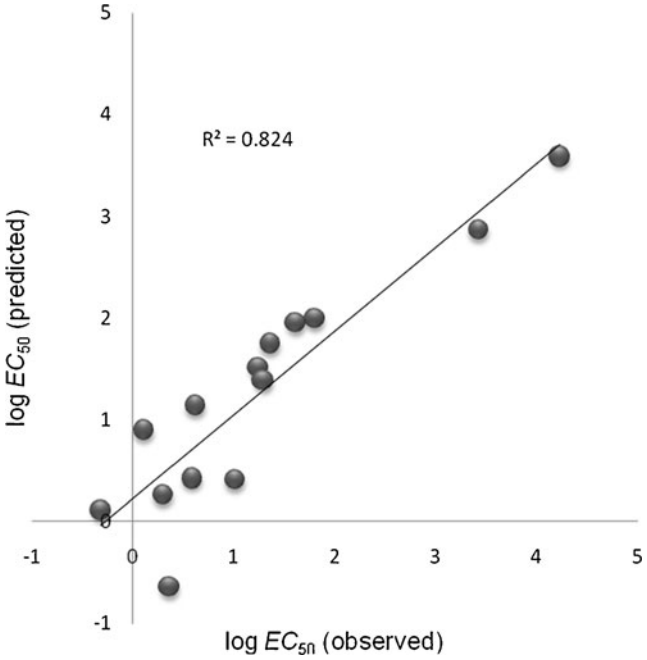

Fig. 3 Observed and predicted values of phototoxicity toward *Daphnia magna* for the set of 14 molecules. Predictions were made using model (11)

model:

$$\begin{aligned}\log EC_{50} &= -24.599 - 8.910E_{\mathrm{HOMO}} - 0.015M + 27.824S \\ &\quad -9.697\chi \\ n &= 14, R^2 = 0.928, R^2{}_{\mathrm{adj}} = 0.896, s = 0.405, \\ F &= 28.878, p_{\mathrm{F}} = 3.8 \times 10^{-5} R^2{}_{\mathrm{CV}} = 0.589.\end{aligned} \quad (13)$$

The t values were then −3.784, −4.206, 3.406, and −4.676 for E_{HOMO}, molecular weight, softness index, and electronegativity, respectively, which indicate that all of these descriptors have significant effects on the photoinduced toxicity. In this four-descriptor model, benzo(*b*)anthracene was an outlier. Removing this molecule significantly improved the model performance ($R_{\mathrm{CV}}{}^2$=0.770).

As judged from the cross-validation coefficients ($R_{\mathrm{CV}}{}^2$), model (13) is statistically superior to model (12) and inferior to model (11).

Set B (phototoxicity of 12 PAHs toward *Scenedemus vacuolatus*)

Figure 4 shows that the vector of the E_{LUMO} descriptor is very close to that of logEC_{50}, implying a significantly positive correlation between them (R=0.91). This means that the energy of the lowest unoccupied molecular orbital has a strong effect on the value of logEC_{50}. The biplot reveals that the descriptors E_{GAP}, octanol/water partition coefficient, chemical hardness, and chemical potential are quite close to logEC_{50} and show positive correlations with logEC_{50} ($0.50 \leq R \leq 0.80$). On the other hand, molecular weight, surface area grid, molecular volume, molar refractivity, polarizablity, and softness index all show negative correlations with logEC_{50} (they are on the opposite side of the circle to it), but they are considered significant descriptors ($R \approx -0.8$). The hydration energy, E_{HOMO}, shape descriptor, and electrophilicity exert only minor effects on the logEC_{50} value considering their low correlation coefficients, as shown in Table 4.

After removing unsuitable descriptors (hydration energy, E_{HOMO}, shape descriptor, and electrophilicity), we again

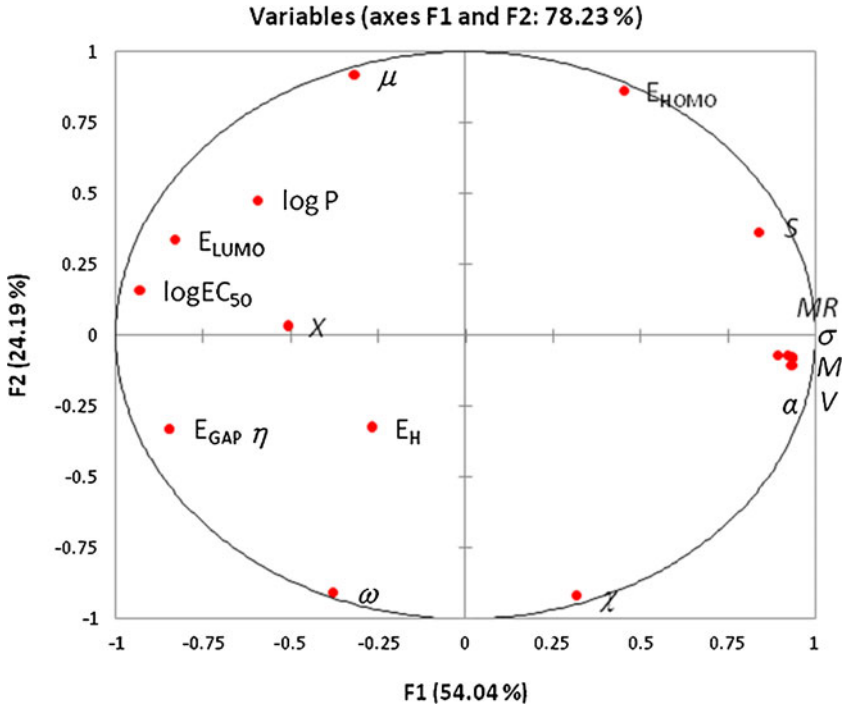

Fig. 4 Correlation circle for phototoxicity toward *Scendemus vacuolatus* and the calculated descriptors

Table 4 Correlation coefficient matrix for phototoxicity toward *Scendemus vacuolatus* and the calculated descriptors

	logEC_{50}	E_H	E_{HOMO}	E_{LUMO}	E_{GAP}	M	log$P$	σ	V	MR	α	X	η	S	μ	χ	ω
logEC_{50}	1																
E_H	0.069	1															
E_{HOMO}	−0.324	−0.215	1														
E_{LUMO}	0.910	−0.094	−0.163	1													
E_{GAP}	0.817	0.075	−0.750	0.775	1												
M	−0.820	−0.302	0.287	−0.671	−0.634	1											
log$P$	0.661	−0.551	0.029	0.735	0.474	−0.604	1										
σ	−0.752	−0.563	0.247	−0.595	−0.557	0.948	−0.368	1									
V	−0.786	−0.469	0.276	−0.624	−0.595	0.982	−0.461	0.991	1								
MR	−0.809	−0.347	0.298	−0.650	−0.627	0.999	−0.564	0.960	0.989	1							
A	−0.810	−0.343	0.297	−0.652	−0.628	0.999	−0.568	0.959	0.988	1.000	1						
X	0.371	0.861	−0.044	0.288	0.221	−0.513	−0.198	−0.755	−0.659	−0.543	−0.540	1					
η	0.817	0.075	−0.750	0.775	1.000	−0.634	0.474	−0.557	−0.595	−0.627	−0.628	0.221	1				
S	−0.787	−0.083	0.773	−0.743	−0.993	0.638	−0.467	0.555	0.597	0.632	0.633	−0.204	−0.993	1			
μ	0.473	−0.237	0.627	0.666	0.045	−0.314	0.602	−0.284	−0.284	−0.288	−0.291	0.195	0.045	−0.003	1		
χ	−0.473	0.237	−0.627	−0.666	−0.045	0.314	−0.602	0.284	0.284	0.288	0.291	−0.195	−0.045	0.003	−1.000	1	
ω	0.228	0.244	−0.993	0.060	0.677	−0.234	−0.103	−0.204	−0.230	−0.249	−0.247	0.030	0.677	−0.707	−0.703	0.703	1

found that the best two-descriptor regression model was based on the molar refractivity and the energy of the lowest unoccupied molecular orbital:

$$\begin{aligned} \log EC_{50} &= 7.303 + 2.213E_{LUMO} - 0.029MR \\ n &= 12, R^2 = 0.910, R^2{}_{adj} = 0.890, s = 0.310, \\ F &= 45.285, p_F = 2.0 \times 10^{-5} \\ R^2{}_{CV} &= 0.837. \end{aligned} \quad (14)$$

The values of t were 5.042 and −2.851 for E_{LUMO} and MR, respectively, which indicate that both descriptors were important, while E_{LUMO} is the most significant. logEC_{50} was again positively correlated with E_{LUMO} and negatively correlated with molar refractivity. The observed values of photoinduced toxicity toward *Scenedemus vacuolatus* are compared numerically in Table 5 to the predictions of the two-descriptor model. The correlation between those predictions and the observed values of photoinduced toxicity toward *Scenedemus vacuolatus* is shown graphically in Fig. 5.

Table 5 Observed and predicted values of phototoxicity toward *Scenedemus vacuolatus* for the set of 12 PAHs. Predictions were made using the two-descriptor model (14)

Molecule name	logEC_{50} (observed)	logEC_{50} (predicted)
Anthracene	2.02	1.89
Benzo[*a*]anthracene	1.06	1.53
Benzo[*a*]pyrene	0.40	0.81
Benzo[*ghi*]perylene	0.86	0.75
Fluoranthene	1.42	1.35
Phenanthrene	3.58	3.30
Pyrene	1.54	1.93
Benzo[*b*]fluoranthene	0.80	0.88
Benzo[*k*]fluoranthene	1.12	0.92
Indeno[1,2,3-*cd*]pyrene	0.36	−0.02
Benzo[*ghi*]fluoranthene	1.01	0.94
2-Phenylnaphthalene	2.56	2.45

The other model that was found to be valuable for this set of molecules was based on three descriptors and took the following form:

$$\begin{aligned} \log EC_{50} &= 4.903 + 1.315E_{GAP} - 0.026MR - 1.838\chi \\ n &= 12, R^2 = 0.921, R^2{}_{adj} = 0.892, s = 0.307, \\ F &= 31.256, p_F = 9.1 \times 10^{-5} \\ R^2{}_{CV} &= 801. \end{aligned} \quad (15)$$

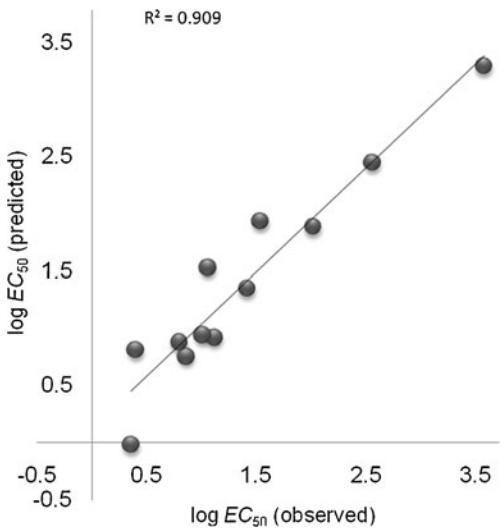

Fig. 5 Observed and predicted values of phototoxicity toward *Scenedemus vacuolatus* for the set of 12 molecules. Predictions were made using model (14)

The values of t were 4.557, −2.505, and −3.322 for E_{GAP}, molar refractivity, and electronegativity, respectively. These values indicate that E_{GAP} has the greatest influence on the photoinduced toxicity of the PAHs. The three-descriptor model was more statistically significant than the model developed by Wang et al. [18], who reported a similar correlation coefficient (R^2=0.921) upon using a large number of molecular descriptors (six descriptors) for the same set of molecules. Based on the cross-validation coefficients (R_{CV}^2), model (15) is statistically inferior to model (14).

Conclusions

Using quantum-chemical descriptors, we demonstrated that very useful multiple linear regression models can be obtained for data on the phototoxicity of polyaromatic hydrocarbons (PAHs; the same set of molecules that were considered previously by Wang et al. [18]) toward *Daphnia magna* and *Scenedesmus vacuolatus* (expressed as logEC_{50}) . E_{HOMO}, E_{LUMO}, E_{GAP}, softness index, electronegativity, molar refractivity (*MR*), and molecular weight were found to be important influences on the photoinduced toxicity of the PAHs toward *Daphnia magna* and *Scenedesmus vacuolatus*. Using only two descriptors, we found that very simple linear models based on E_{LUMO} and *MR* perform well for the two sets of molecules. For phototoxicity toward *Scenedesmus vacuolatus*, a remarkable three-descriptor regression model based on E_{GAP}, electronegativity, and molar refractivity was constructed. Our various QSAR models were generally more statistically significant than those reported in [18], which were based on as many as six descriptors. The resulting models demonstrate the potential benefits of incorporating quantum-chemical descriptors into QSAR models for predicting the phototoxicity of PAHs.

References

1. Mihelcic JR, Luthy RG (1988) Appl Environ Microbiol 54:1182–1187
2. Newsted JL, Giesy JP (1987) Environ Toxicol Chem 6:445–461
3. Mekenyan OG, Ankley GT, Veith GD, Call DJ (1994) Chemosphere 28:567–582
4. Huovinen PS, Soimasuo MR, Oikari AOJ (2001) Chemosphere 45:683–691
5. Huang XD, Dixon DG, Greenberg BM (1993) Environ Toxicol Chem 12:1067–1077
6. Ren L, Zeiler LF, Dixon DG, Greenberg BM (1996) Ecotox Environ Saf 33:73–80
7. Little EE, Cleveland L, Calfee R, Barron MG (2000) Environ Toxicol Chem 19:926–932
8. Foote CS (1991) Photochem Photobiol 54:659–659
9. Cronin MTD (2002) ATLA 30:81–84
10. Netzeva TI, Dearden JC, Edwards R, Worgan ADP, Cronin MTD (2004) J Chem Inf Comput Sci 44:258–265
11. Karelson M, Labanov VS, Katritzky AR (1996) Chem Rev 96:1027–1043
12. Veith GD, Mekenyan OG, Ankley GT, Call DJ (1995) Chemosphere 30:2129–2142
13. Dong S, Fu PP, Shirsat RN, Hwang HM, Leszczynski J, Yu H (2002) Chem Res Toxicol 15:400–407
14. Grote M, Schüürmann G, Altenburger R (2005) Environ Sci Technol 39:4141–4149
15. Eroglu E, Türkmen H (2007) J Mol Graph Model 26:701–708
16. Gu CG, Jiang X, Ju XH, Gong XD, Wang F, Bian YR, Sun C (2009) Ecotoxicol Environ Saf 72:60–70
17. Pasha FA, Srivastava HK, Singh PP (2005) Bioorg Med Chem 13:6823–6829
18. Wang Y, Chen J, Li F, Qin H, Qiao X, Hao C (2009) Chemosphere 76:999–1005
19. Parr RG, Donnelly RA, Levy M, Palke WE (1978) J Chem Phys 68:3801–3807
20. Parr RG, Pearson RG (1983) J Am Chem Soc 105:7512–7516
21. Parr RG, Szentpaly LV, Liu S (1999) J Am Chem Soc 121:1922–1924
22. Al-Fahemi JH, Cooper DL, Allan NL (2005) Theochem 727:57–61
23. Ribeiro FAL, Ferreira MMC (2005) Theochem 719:191–200
24. SPSS Inc. (1999) SPSS Base 13.0 applications guide. SPSS Inc., Chicago