ORIGINAL PAPER

QSPR/QSAR solely based on molecular surface electrostatic potentials for benzenoid hydrocarbons

Raouf Ghavami[1] · **Bakhtyar Sepehri**[1]

Received: 15 May 2015 / Accepted: 8 October 2015 / Published online: 27 October 2015

Abstract Benzenoid hydrocarbons are a group of the most important π-electron systems having the attention of both experimental and theoretical chemists for the last 100 years. In the present study, based on the general interaction properties function (GIPF) family descriptors, significant one- or two-parametric quantitative structure–property (activity) relationship models were developed for the prediction of properties/activities of benzenoids hydrocarbons. All descriptors were computed in density functional theory (DFT) at the B3LYP/STO-3G level of theory in Gaussian98 software. A large number of physico-chemical properties and two biological activities (e.g. bio-concentration factor and photo-induced toxicity) of these compounds were investigated by using multiple linear regressions. All created models were interpreted in term of selected descriptors. R^2 and R^2_{cv} values of all models are respectively between 0.665–0.994 and 0.609–0.990 for the whole dataset of each property/activity. Maximum R^2 for Y-randomization ($R^2_{\max}$) test and its cross-validation (R^2_{cv}, max) are between 0.098–0.485 and 0.002–0.357, respectively.

Keywords Benzenoid hydrocarbons · QSPR · QSAR · Molecular surface electrostatic potentials · GIPF approach · Local ionization energy

Introduction

Benzenoid hydrocarbons are ubiquitous and are found in all classes of natural products, pharmaceuticals, and materials. Benzenoid hydrocarbons containing two or more fused benzene rings are classes of organic pollutants that are produced during the incomplete burning of coal, oil, gas, wood, garbage or other organic substances that resulting from human activities. These compounds are widely found in the environment and foods such as vegetables [1–4]. Since, some common benzenoid hydrocarbons have been known to be potent carcinogens, this contaminant class is generally regarded as having high priority for environmental pollution regulation and in ecological risk assessment of industrial effluent discharges. In relation to water, most hydrophobic benzenoids will typically absorb strongly to particles and become generally more resistant to bacterial degradation [5]. The most concerning matter on benzenoid hydrocarbons is that they have shown to be highly carcinogenic in laboratory animals, and it's also involved in different types of human cancers, mainly breast, lung, and colon cancer. The metabolic activations of these compounds in mammalian cell to dioepoxides cause errors in DNA replication and mutation, which initiates the carcinogenic process [6]. Some benzenoids have chemical stability and spermatogenetic and mutagenic effects [3]. One of the most successful approaches to the prediction of physico-chemical properties and biological activities of compounds is quantitative structure–property/activity relationships modeling (QSPRs/QSARs) [7]. QSPRs/QSARs are mathematical models that attempt to correlate the molecular structure of compounds to their biological, chemical, and physical properties. The main steps comprising this method are: data collection, molecular geometry optimization, molecular descriptors generation, descriptor selection,

Electronic supplementary material The online version of this article (doi:10.1007/s13738-015-0761-2) contains supplementary material, which is available to authorized users.

✉ Raouf Ghavami
rghavami2000@yahoo.com; r.ghavami@uok.ac.ir

[1] Department of Chemistry, Faculty of Science, University of Kurdistan, P. O. Box 416, Sanandaj, Iran

model development and finally model performance evaluation [8, 9]. The main problems encountered in this kind of research are still the description of the molecular structure using appropriate molecular descriptors and selection of suitable modeling methods [10]. Currently, many types of molecular descriptors such as topological indices and quantum chemical parameters have been proposed to describe the structural feature of molecules [11–13]. Ferreira modeled seven properties and two biological activities of polycyclic aromatic hydrocarbons (PAHs) by using partial least squares (PLS) regression and electronic descriptors such as energy of HOMO and LUMO orbitals, topological descriptors and geometric descriptors such as molecular volume and surface area and heat of formation [14]. Lu et al. have predicted photolysis half-lives of polycyclic aromatic hydrocarbons using PLS and computed quantum chemical descriptors from Gaussian03 [15]. In another research, Nikolić created linear relationship between polarographic half-wave reduction potentials of benzenoid hydrocarbons and electron affinities and E_{LUMO} as descriptors [16]. In addition, aqueous solubility of PAHs have been predicted using GA-SVM and GAs-RBFNs and molecular connectivity indices [3].

It has long been recognized that non-covalent interactions are predominantly electrostatic in nature. Politzer et al. have shown that a variety of condensed phase macroscopic properties that depend on non-covalent interactions can be expressed analytically in terms of statistically defined quantities that characterize molecular surface electrostatic potentials and average local ionization energy and named their model, general interaction properties function (GIPF) [17, 18]. Prediction properties such as $\mathrm{pK_a}$ that includes charge transfer need molecular surface average local ionization energy descriptors [19]. Molecular volumes consider polarizability effects because there exists a linear relationship between them [20]. Some properties depend on interactions that are non-covalent such as solubility in water, boiling point and so on; therefore, these descriptors can be used for predicting these properties. In this research, we attempted to develop QSPR/QSAR models for predicting the physico-chemical properties and biological activities of benzenoid hydrocarbons using MLR and GIPF family descriptors.

Materials and methods

Data for benzenoids

All experimentally determined physico-chemical properties and biological activities of benzenoid hydrocarbons including the boiling point at normal pressure (BP, in °C), retention index (RI) refers to the reversed-phase liquid chromatography on polymer, aqueous solubility (log $S_{\mathrm{w,l}}$), lipophilicity or *n*-octanol/water partition coefficient (log K_{OW}), *n*-octanol/air partition coefficient (log K_{OA}), soil sorption (log K_{OC}), Henry's law constant (log *H*), bioconcentration factor (log BFC), photo-induced toxicity (log $1/\mathrm{LT}_{50}$), polarographic half-wave reduction potentials ($E_{1/2}$), heat (enthalpy) of formation (ΔH_{f}), photolysis half-lives (log $t_{1/2}$) and the molecular resonance energy (RE) were taken from the literature [14–16, 21–26]. The experimental values of these properties and activities are listed in Table S1 of the supplementary information.

Computer hardware and software

All calculations were performed on a 2.5 GHz Intel® CoreTM2 Quad Q 8300 CPU with 2 GB of RAM using all four available cores under Windows XP operating system. The ISIS/Draw version 2.3 software was used for drawing the molecular structures [27]. Molecular modeling and geometry optimization were employed by HyperChem (version 7.1, HyperCube, Inc.) [28]. Gaussian98 program [29] was used to re-optimize the molecular structure. SPSS software (version 16.0, SPSS, Inc.) http://www.spss.com/ was used for elimination stepwise selection MLR analysis and other computations were performed in the MATLAB (version 7.0, Math Works, Inc.) environment.

Molecular descriptors generation and calculation

First we created and optimized 48 benzenoid hydrocarbon molecules in HyperChem 7.1 using AM1 method. Then re-optimization were implemented in Gaussian98 software at B3LYP/STO-3G level. Next, these optimized geometries were used to compute the electrostatic potential $V(r)$ on the molecular surfaces that is defined by the 0.001 au contour of the electron density $\rho(r)$. Molecular surface electrostatic potentials were computed at B3LYP/6-31G* by Gaussian98 software. The grid control option was set to "cube = 100". Thus, for each molecule, Molecular surface electrostatic potentials were computed at approximately 100^3 points. Then we used the WFA (wave function analysis) statistical analysis program to compute molecular surface electrostatic potential and average local ionization energy descriptor using the produced CUBE file with Gaussian98 [29–32]. The electrostatic potential $V_{\mathrm{s}}(r)$ in the space around a molecule that is created by its nuclei and electrons is defined by Eq. (1):

$$V_{\mathrm{s}}(r) = \sum_{A} \frac{Z_A}{|R_a - r|} - \int \frac{\rho(r')\mathrm{d}r'}{|r' - r|} \qquad (1)$$

where Z_A is the charge on nucleus A, located at R_A. The first term on the right side of Eq. (1) is the nuclear contribution to $V(r)$ which is positive, the second term is due to the electrons and is negative [33, 34]. The average local ionization energy, $\overline{I}(r)$, is defined by Eq. (2):

$$\overline{I}(r)=\frac{\sum_i \rho_i(r)|\varepsilon_i|}{\rho(r)} \tag{2}$$

$\rho_i(r)$ is the electronic density of the molecular orbital at the point r, ε_i is its orbital energy and $\rho(r)$ is the electronic density function. We interpret $\overline{I}(r)$ as the required energy, on average, to remove an electron from a point r in the space of an atom or a molecule [31, 35]. $V_{\mathrm{S}}(r)$ is effective for non-covalent interactions, which are largely electrostatic in nature, while $\overline{I}_{\mathrm{S}}(r)$ is more suitable when there is transfer of charge (electron pair donor–electron pair acceptor interaction) that is one of the forces responsible for separation of compounds in chromatography [31, 33, 36]. It might seem that $V_{\mathrm{S}}(r)$ could also predict sites for electrophilic and nucleophilic bond-forming attack, by means of its most negative and positive regions. However $V_{\mathrm{S}}(r)$ is not consistently reliable in this respect, because the regions of most negative $V_{\mathrm{S}}(r)$ do not always correspond to the sites where the most reactive electrons are located. For example, the most negative $V_{\mathrm{S}}(r)$ in benzene derivatives such as aniline, phenol, fluoro- and chlorobenzene, and nitrobenzene are associated with the substituents, whereas electrophilic reaction occurs on the rings. In contrast, $\overline{I}_{\mathrm{S}}(r)$ correctly predicts the *ortho*/*para*- or *meta* directing effects of the substituents, as well as their activation or deactivation of the ring [31].

Politzer et al. developed an approach which can be summarized as Eq. (3) and named it, general interaction properties function (GIPF) [17, 31]:

$$\text{Property}=f\left(V_{\mathrm{mv}},A_{\mathrm{s}}^{\mathrm{tot}},A_{\mathrm{s}}^{+},A_{\mathrm{s}}^{-},V_{\mathrm{s,max}},V_{\mathrm{s,min}},\overline{V}_{\mathrm{s}},\overline{V}_{\mathrm{s}}^{+},\overline{V}_{\mathrm{s}}^{-},\pi^{\mathrm{tot}},\delta_{\mathrm{tot}}^{2},\delta_{+}^{2},\delta_{-}^{2},\vartheta,\overline{I}_{\mathrm{s,max}},\overline{I}_{\mathrm{s,min}},\overline{\overline{I}}_{s},\delta_{\overline{I}_{\mathrm{s}}}^{2},\pi_{\overline{I}_{\mathrm{s}}}\right) \tag{3}$$

In this equation, V_{mv} is the molecular volume and $A_{\mathrm{s}}^{\mathrm{tot}}$, A_{s}^{+}, A_{s}^{-} are total surface area and the surface area over which $V_{\mathrm{s}}(r)$ is positive and negative, respectively. $V_{\mathrm{s,max}}, V_{\mathrm{s,min}}$, are maxima and minima of electrostatic potential on the molecular surface and $\overline{V}_{\mathrm{s}}$, $\overline{V}_{\mathrm{s}}^{+}$ and $\overline{V}_{\mathrm{s}}^{-}$ are respectively, the overall average potentials and the average of positive and negative potentials are computed as:

$$\overline{V}_s=\frac{1}{t}\sum_{i=1}^{t}V_s(r_i),\overline{V}_s^{+}=\frac{1}{m}\sum_{j=1}^{m}V_s^{+}(r_j),\bar{V}_s^{-}=\frac{1}{n}\sum_{k=1}^{n}\overline{V}_s^{-}(r_k) \tag{4}$$

where π^{tot} is the average deviation of overall potentials and is computed as:

$$\pi=\frac{1}{t}\sum_{i=1}^{t}\left|V_s(r_i)-\overline{V}_s\right| \tag{5}$$

π is interpreted as an indicator of internal charge separation, which is present even in molecules having zero dipole moment due to symmetry, e.g. para-dinitrobenzene and boron trifluoride. $\delta_{\mathrm{tot}}^{2}, \delta_{+}^{2}, \delta_{-}^{2}$ are respectively total, positive and negative variances and are computed as:

$$\delta_{tot}^{2}=\delta_{+}^{2}+\delta_{-}^{2}=\frac{1}{m}\sum_{j=1}^{m}\left[V_s^{+}(r_j)-\overline{V}_s^{+}\right]^2+\frac{1}{n}\sum_{k=1}^{n}\left[V_s^{-}(r_k)-\overline{V}_s^{-}\right]^2 \tag{6}$$

where ν is electrostatic balance parameter and is computed as [31, 33]:

$$\nu=\frac{\delta_{+}^{2}\delta_{-}^{2}}{\left[\delta_{+}^{2}+\delta_{-}^{2}\right]^2} \tag{7}$$

In the summations above, t is the total number of points on the surface grid, m and n are the numbers of points at which $V(r)$ is positive and negative, respectively. The features of $\overline{I}(r)$ could be characterized analogously to those of $V(r)$, that are extrema $\overline{I}_{\mathrm{s,max}}, \overline{I}_{\mathrm{s,min}}$, its average magnitude $\overline{\overline{I}}_{\mathrm{s}}$, average deviation ($\pi_{\overline{I}_{\mathrm{s}}}$), and variance ($\delta_{\overline{I}_{\mathrm{s}}}^{2}$)—keeping in mind that $\overline{I}(r)$ only takes positive values [7, 31].

Descriptors selection

The GIPF family descriptors consisted of 14 surface electrostatic potential, five average local ionization energies, Bader molecular volumes were calculated which have been listed in Table S2. Then 16 combinations of descriptors were calculated from original GIPF descriptors that have been listed in Table S3. Therefore, we calculated 35 descriptors that were used to build QSAR/QSPR models. In order to minimize the information overlapping in descriptors and to reduce the number of descriptors required in regression equation, the concept of non-redundant descriptors (NRD) were used [37]. That is, when two descriptors are correlated by a linear correlation coefficient value greater than 0.85, both descriptors are correlated with a dependent variables, the better correlation is used for the actual analysis, discarding the descriptors with a lower correlation. This objective-based feature selection left reduced and predictive descriptors for the studied compounds. Using these criteria for each physico-chemical property or biological activity, z descriptors out of 35 original descriptors were eliminated and 35-z

Table 1 Best models for the studied properties/activities and statistical parameters of benzenoids

Property/activity	n	Descriptors	R^2	R^2_{adj}	RMSE	F	R^2_{cv}	RMSE$_{\mathrm{cv}}$	R^2_{max}	R^2_{cv},max
BP	23	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^+$	0.986	0.985	15.3	1462.1	0.98	18.15	0.103	0.035
RI	33	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^+, \bar{I}_{\mathrm{s,max}}$	0.942	0.938	0.24	242.8	0.929	0.27	0.118	0.002
log $\mathrm{S}_{\mathrm{w,l}}$	15	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^+$	0.932	0.926	0.52	177.2	0.914	0.59	0.177	0.085
log K_{OW}	18	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^+$	0.989	0.988	0.14	1453.8	0.987	0.16	0.098	0.019
log K_{OA}	9	$A_\mathrm{s}^+ \delta_+^2$	0.988	0.986	0.3	562.7	0.982	0.37	0.477	0.303
log K_{OC}	9	$A_\mathrm{s}^{\mathrm{tot}}$	0.994	0.993	0.13	1176.8	0.99	0.18	0.46	0.357
log H	8	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^-$	0.954	0.946	0.23	124.8	0.924	0.3	0.458	0.255
log BCF	11	$\overline{V}_\mathrm{s}^+$	0.887	0.875	0.24	70.9	0.844	0.29	0.27	0.096
log 1/LT_{50}	9	$\delta_{\mathrm{tot}}^2, A_s^+$	0.802	0.752	0.14	16.2	0.62	0.2	0.479	0.261
$E_{1/2}$	27	$\delta_-^2, V_{\mathrm{s,max}}$	0.665	0.637	0.18	23.9	0.609	0.19	0.122	0.034
ΔH_f	20	A_s^-	0.985	0.984	8.15	1164	0.981	9.1	0.257	0.153
log $t_{1/2}$	7	$\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^-$	0.904	0.885	0.13	47.1	0.851	0.16	0.485	0.344
RE	20	$(\mathrm{A}_\mathrm{s}^- \overline{V}_\mathrm{s}^-)^2$	0.823	0.813	0.33	83.8	0.794	0.36	0.131	0.058

descriptors remained. In GIPF approach, properties/activities of molecules are related to a few number of descriptors; therefore a variable reduction technique is needed. In this study, the most important variables are selected by elimination stepwise selection procedure, which combines the forward selection and backward elimination approaches. Initially, we consider the descriptive variable, which has the highest correlation with the response. If the inclusion of this variable results in a significant improvement of the regression model, it is retained and the selection continues. In the next step, the variable that gives the most significant decrease of the regression sum of squares is added to the model. After each forward selection step a backward elimination step is performed. In this step, a partial F test for the variables, already presented in the equation, is carried out. If a variable does not contribute significantly in the building of the regression model, then it will be removed. The procedure stops at the condition that no variables fulfill the requirements anymore to be removed or entered. After this selection procedure, classical MLR can be applied on the retained variables to build a predictive model [7, 38, 39].

Multiple linear regression (MLR)

An MLR model assumes that there is a linear relationship between the molecular descriptors of a compound, which is usually expressed as a feature vector X (where each entry indicates a descriptor), and its target property, y. An MLR model can be described using the following equation:

$$y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + \varepsilon_i, i = 1, 2, \ldots, n \tag{8}$$

where $\{X_{i1},\ldots,X_{ik}\}$ are molecular descriptors, β_0 is the regression model constant, β_1–β_k are the coefficients corresponding to the descriptors X_{i1} to X_{ik} and y is dependent variable [39]. The values for β_0–β_k are chosen by minimizing the sum of squared vertical distances of the points from the hyper plane so as to give the best prediction of y from X. The molecular descriptors should be mathematically independent (orthogonal) to each other and the number of compounds in the training set should exceed the number of molecular descriptors by at least a factor of 4 [38, 40]. In this research, statistical parameters including R^2, squared correlation coefficient, R^2_{adj}, adjusted squared correlation coefficient, RMSE, root mean squared error; REP, relative error prediction and F, F test (Fischer's value) were calculated for each model:

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \tag{9}$$

where SSE and SST are the sum of squared errors and the total sum of squares, respectively; and calculated as:

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \cdots - \widehat{\beta}_k x_{ik})^2 \tag{10}$$

$$\mathrm{SST} = \sum_{i=1}^{n} (y_i - \overline{y}_i)^2 \tag{11}$$

and other parameters were calculated as:

$$R^2_{\mathrm{adj}} = 1 - \left(1 - R^2\right) \frac{(n-1)}{(n-k-1)} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2} \tag{13}$$

$$\mathrm{REP}\ (\%) = \frac{100}{\hat{y}_i}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{14}$$

$$F = \frac{(n-k-1)R^2}{k(1-R^2)} \tag{15}$$

In the mentioned equations above, n is the number of compounds, k is the number of variables, y_i is the experimental property/activity, $\hat{y}_i$ is the average of experimental property/activity and $\hat{y}_i$ is the calculated property/activity from QSPR/QSAR model [41, 42]. These parameters have been listed in Table 1 for all models.

Model prediction-validation

Model validation is a critical component of QSPR/QSAR development. A number of procedures has been established to determine the quality of models. In this research, a leave-one-out cross-validation (LOO-CV) and Y-randomization test are used to validate the predictive ability and check the statistical significance of the developed models.

Leave-one-out cross-validation (LOO-CV)

In LOO-CV, each time a molecule was removed and variable selection was performed for remaining molecules and a new model was built. This new model was used to predict the property (or activity) of removed molecules and this process was repeated for n times that n is the number of molecules that were used in the model building. Finally, the cross-validated squared correlation coefficient, R^2_{CV} and root mean square error in cross-validation, $\mathrm{RMSE}_{\mathrm{CV}}$, were for each model calculated as:

$$R^2_{\mathrm{CV}} = 1 - \frac{\sum_{i=1}^{N}\left(y_{\mathrm{pred,i}} - y_{\mathrm{obs,i}}\right)^2}{\sum_{i=1}^{N}\left(y_{\mathrm{obs,i}} - \bar{y}_{\mathrm{obs}}\right)^2} = 1 - \frac{PRESS}{\sum_{i=1}^{N}\left(y_{\mathrm{obs,i}} - \bar{y}_{\mathrm{obs}}\right)^2} \tag{16}$$

$$\mathrm{RMSE}_{\mathrm{CV}} = \sqrt{\frac{\sum_{i=1}^{N}(y_{\mathrm{pred,i}} - y_{\mathrm{obs,i}})^2}{N}} \tag{17}$$

where n is the number of training patters, $y_{\mathrm{obs,i}}$ and $y_{\mathrm{pred,i}}$ are the experimental, and predicted property/activity of the left-out benzenoid hydrocarbon i, respectively and $\bar{y}_{Iobs}$ is the average of experimental property/activity of molecules [43–46].

Y-randomization test

The Y-randomization of response is another important validation approach that is widely used to establish model robustness. In this test, dependent variable is reordered randomly and a new model is built. The procedure was repeated 100 times and the best model that has the maximum R^2 (R^2_{max}) and maximum cross-validated R^2 ($R^2_{cv,max}$) was selected. Small values of R^2_{max} and $R^2_{cv,max}$ demonstrate that QSAR/QSPR model has not been obtained by chance [47–50]. These parameters have been listed in Table 1 for all models.

Results and discussion

The best models were obtained by elimination stepwise selection regression algorithm and the statistical parameters for the models and their cross-validation tests were summarized in Table 1. It is interesting to note that for these data sets the combination descriptors $\mathrm{A}_\mathrm{s}^-\overline{V}_s^+$ and $A_s^+\overline{V}_s^+$ (as obviously, is demonstrated that positive and negative electrostatic potential regions of benzenoid hydrocarbons interact with each other or with solvent molecules) gives superior prediction power in the QSPR/QSAR models for several studied properties/activities. Also, among the thirteen obtained models, ten of them are mono-parametric and the rest are bi-parametric models. Finally, we predicted the values of all properties/activities for all benzenoids by creating the best models (selected in this paper) which have been listed in Table S4.

Boiling point (BP, °C)

The observed data of 23 benzenoids that have been listed in Table S1 were used to construct the QSPR model. After descriptors selection step, the following equation with a combinatorial descriptor was built:

$$\mathrm{BP} = -61.104(\pm 14.197) + 0.433(\pm 0.011)\mathrm{A}_\mathrm{s}^-\overline{V}_s^+ \tag{18}$$

According to this equation, if a benzenoid has more points with negative electrostatic potential and also more positive average potential in its surface so has more electrostatic attraction between its molecules and its boiling point increase. For benzenoids with more than two rings, balance parameter value (υ) is near to its maximum value that is 0.25. This means that benzenoids can interact up to a similar extent (whether strongly or weakly) through its both positive and negative electrostatic potential regions [51]. Although for benzenoids $A_s^+ \cong A_s^-$, but positive electrostatic points on molecular surface do not centralize. Thus a region with negative and positive electrostatic potential points cannot coincide as can be seen in Figs. 1 and 2, so in the model $\overline{V}_s^+$ has been selected rather than A_s^+ is chosen. The resulting plot for the mono-parametric model is shown in supplementary information Fig. SA.

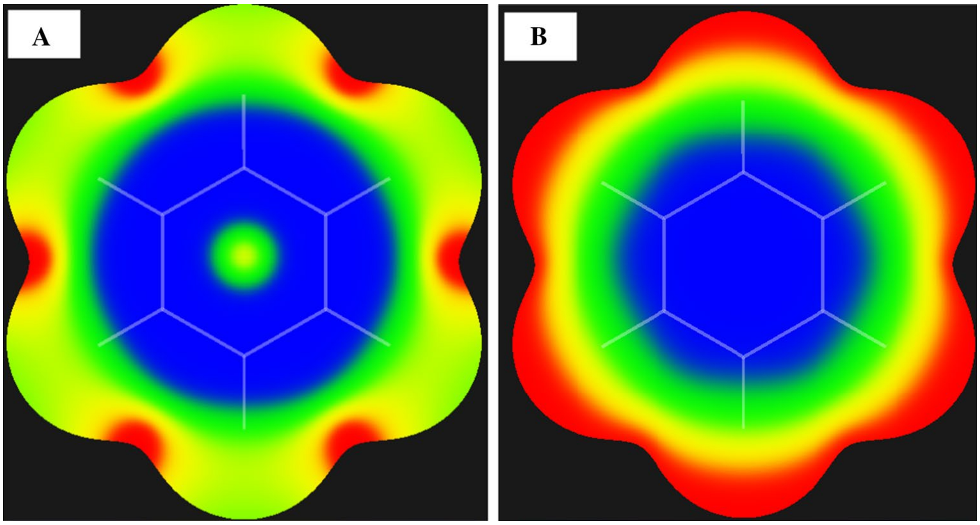

Fig. 1 **a** Calculated B3LYP/STO-3G ionization energy on molecular surface of benzene. Ionization ranges in eV/mol: *red* more than 12.4640, *yellow* between 12.4640 and 11.2937, *green* between 11.2937 and 10.1234, *blue* smaller than 10.1234, **b** Calculated B3LYP/6-31G* electrostatic potential molecular surface of benzene. Electrostatic potential ranges in Kcal/mol: *red* more than 4.6413, *yellow* between 4.6413 and −2.9282, *green* between −2.9282 and −10.4977, *blue* more negative than −10.4977

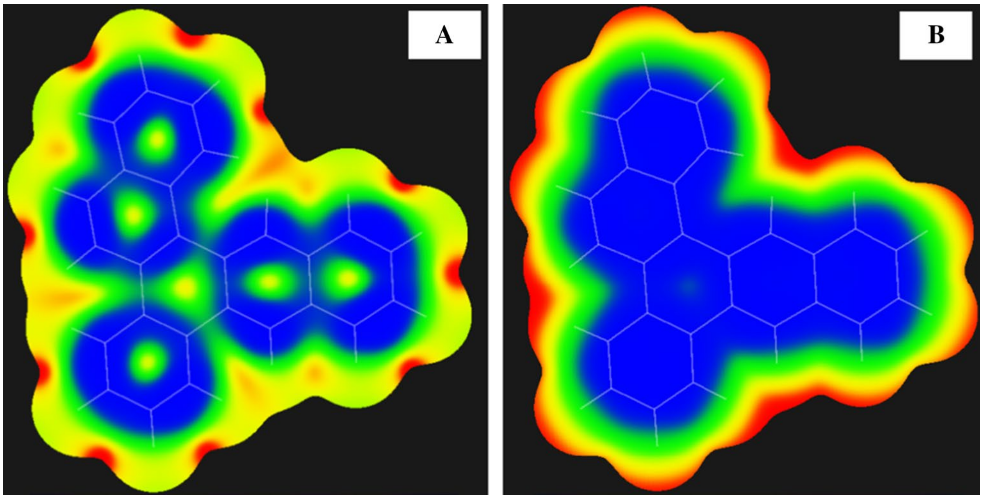

Fig. 2 **a** Calculated B3LYP/STO-3G ionization energy on molecular surface of dibenz[*a*,*n*]triphenylene. Ionization ranges in eV/mol: *red* more than 12.5615, *yellow* between 12.5615 and 11.2963, *green* between 11.2963 and 10.0310, *blue* smaller than 10.0310. **b** Calculated B3LYP/6-31G* electrostatic potential molecular surface of dibenz[*a*,*n*]triphenylene. Electrostatic potential ranges in Kcal/mol: *red* more than 10.8232, *yellow* between 10.8232 and 1.9934, *green* between 1.9934 and −6.8364, *blue* more negative than −6.8364

Retention index (RI)

For 33 benzenoids, retention index data values were available that have been listed in Table 1 and after descriptors selection steps, the following model was created:

$$RI = 77.769(\pm 14.438) + 0.004323(\pm 0.000216)A_s^- \bar{V}_s^+ - 5.7459(\pm 1.0638)\bar{I}_{s,max} \qquad (19)$$

These descriptors have no collinearity ($R^2 = 0.3972$). $A_s^+\overline{V}_s^+$ demonstrate that positive and negative electrostatic potential regions of benzenoids and stationary phase attract each other and this attraction is responsible for separation of benzenoids. Negative coefficient of another descriptor shows oppositional effect of this descriptor in the separation mechanism. The R^2_{adj} for the model changed from 0.882 to 0.938 when $\overline{I}_{s,max}$ was added to model. For dibenzo[*a*,*h*]pyrene molecule residual is more than twice the standard deviation of residual of retention index so this molecule is detected as an outlier. For the retention index of benzenoids, a good compatibility for the biparametric regressions is observed in supplementary information Fig. SB and also the resulted model suggests a mechanism for separation of benzenoids.

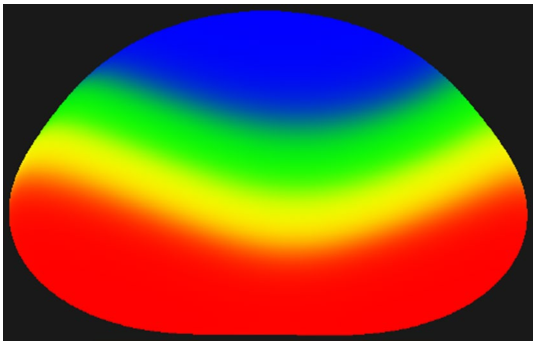

Fig. 3 Calculated B3LYP/STO-3G electrostatic potential molecular surface of water. Electrostatic potential ranges in Kcal/mol: *red* more than 22.3931, *yellow* between 22.3931, and 2.1572, *green* between 2.1572 and −18.0787, *blue* more negative than −18.0787

Water solubility (log $S_{w,l}$)

For the water solubility of benzenoids (log $S_{w,l}$ range −3.85 to 3.28), the few available data allow only a moderate agreement between experimental and calculated values, the best-obtained model is:

$$\log S_{w,l} = 5.5908(\pm 0.56275) - 0.00664(\pm 0.000499)A_s^- \overline{V}_s^+ \quad (20)$$

In this equation, negative coefficient for descriptors is due to the repulsion forces between regions of water and benzenoids molecules that have electrostatic potential with the same sign. In water and benzenoids, regions with negative electrostatic potential exist on oxygen atom and benzene rings (see Figs. 1, 2, 3) that cause repulsion. In addition, positive electrostatic potential on hydrogen's atoms create repulsion forces. Since positive electrostatic points on molecular surface of benzenoids have not been centralized (Figs. 1, 2) so in the model $\overline{V}_s^+$ have been selected instead of A_s^+. For benzenoids, charge separation is low rather than water (see Table S1; Fig. 3), so charge centers are not separated and repulsion forces between them overcome their attraction forces. For Eq. (20) naphthacene is outlier and when this molecule is removed; R^2 increases to 0.952. Fig. SC of supplementary information presents the resulting mono-parametric plot for log $S_{w,l}$ of benzenoids. For the water solubility of benzenoids rather than boiling point, a few available data result in weaker agreement between experimental and calculated values, as can be seen in Fig. SC.

n-Octanol/water partition coefficient (log K_{OW})

The data set for lipophilicity or *n*-octanol/water partition coefficient (log K_{OW} range 2.23–7.19) were included 18 benzenoids which were modeled by mono-parametric equation:

$$\log K_{OW} = 0.27607(\pm 0.1475) + 0.004805(\pm 0.000126)A_s^- \overline{V}_s^+ \quad (21)$$

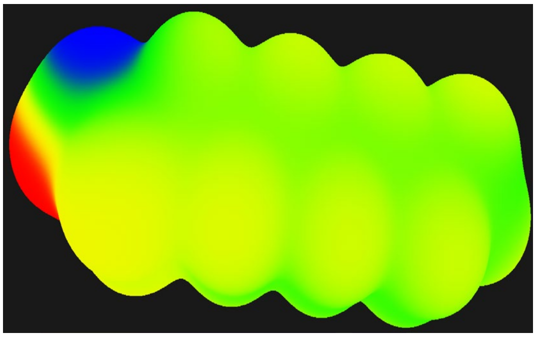

Fig. 4 Calculated B3LYP/STO-3G electrostatic potential molecular surface of octanol. Electrostatic potential ranges in Kcal/mol: *red* more than 20.9836, *yellow* between 20.9836 and 1.8806, *green* between 1.8806 and −17.2223, *blue* more negative than −17.2223

Graph of electrostatic potential on surface of *n*-octanol shows that alkyl section has positive electrostatic potential between 20.9836 and 1.8806 kcal/mol that establish attraction force with negative region of benzenoids (Figs. 1, 2, 4). On the other hand, similar interactions between hydrogen atoms in benzenoids and oxygen atom in *n*-octanol exist. Selection of A_s^- demonstrates that there are many points with positive electrostatic potential in the surface of receptor (here *n*-octanol) that interact with negative points of benzenoids (see Fig. 4). This equation shows interactions between many points with positive and negative electrostatic potential dissolves benzenoids in *n*-octanol. By removing tryphenylene and dibenz[*a*,*h*]anthracene that are outlier, R^2 increases from 0.990 to 0.996. As seen from supplementary information Fig. SD, there exist very good agreements between experimental and calculated values.

Correlation between log K_{OW} and log $S_{w,l}$

As can be seen from comparing Eqs. (20) and (21) for predicting of log $S_{w,l}$ and log K_{OW}, the combination selected descriptor is common to take opposite sign. This means that the lipophilicity and solubility of benzenoids act against each other [52]. According to Table S1 there are thirteen compounds in common between these two properties and the general relationship between log $S_{w,l}$ and log K_{OW} as variables were analyzed using regression analysis as following:

$$\log S_{w,l} = 6.258\ (\pm 0.550) - 1.451\ (\pm 0.100)\log K_{OW} \quad (22)$$

$n = 13$, $R^2 = 0.950$, $R^2_{\text{adj}} = 0.945$, RMSE $= 0.441$, $F = 208.7$, $R^2_{\text{CV}} = 0.937$, $\text{RMSE}_{\text{CV}} = 0.494$

The values of R^2_{adj} for this relationship and Eq. (21) are 0.945 and 0.926, respectively which have no significant difference and demonstrate that the predictive power of $\text{A}_\text{s}^- \overline{V}_\text{s}^+$ is almost the same as lipophilicity (log K_{OW}) that is an experimental descriptor. On the other hand, this is a very good reason that proves that the solubility and the lipophilicity of compounds are related to the interactions between solute and solvent molecules that are electrostatic in nature.

n-Octanol/air partition coefficient (log K_{OA})

The octanol–air partition coefficient is a key descriptor of chemical partitioning between the atmosphere and other environmental organic phases such as soil and vegetation [53, 54]. K_{OA} data values were available for nine compounds in the range between 5.13 and 13.91 log units which were used in this section and the best obtained model is:

$$\log K_{\text{OA}} = -1.9464(\pm 0.49588) + 0.005708(\pm 0.000241)A_\text{s}^+\delta_+^2 \tag{23}$$

The square correlation coefficient between $A_\text{s}^+\delta_+^2$ and $\text{A}_\text{s}^- \overline{V}_\text{s}^+$ is 0.9731, so these descriptors are collinear. Also by increasing the number of rings in benzenoids A_s^+, δ_+^2 and thus $A_\text{s}^+\delta_+^2$ increase which indicates that the molecules become more lipophilic and more soluble in *n*-octanol. We mentioned interaction mechanism between benzenoids and *n*-octanol in *n*-octanol/water partition coefficient (log K_{OW}) section. Fig. SE of supplementary information presents mono-parametric regression according to Eq. (23) and there are very good agreements existing between calculated and experimental values.

Soil sorption (log K_{OC})

Nine benzenoids had soil sorption data and the selected descriptor was a total surface area that resulted to the following equation:

$$\log K_{\text{oc}} = -2.6829(\pm 0.22493) + 0.032824(\pm 0.000957)A_\text{s}^{\text{tot}} \tag{24}$$

A_s^{tot} and $A_\text{s}^+\delta_+^2$ descriptors are collinear ($R^2 = 0.9123$) and R^2 between A_s^{tot} and $\text{A}_\text{s}^- \overline{V}_\text{s}^+$ is 0.9919 which demonstrates that the mentioned interaction mechanism above (see "*n*-octanol/water partition coefficient (log K_{OW})" section) exists between benzenoids and *n*-octanol. Fig. SF of supplementary information presents good agreements between calculated and experimental values. These results demonstrate that A_s^{tot} can completely describe the change in log K_{oc}.

Henry's law constant (log *H*)

Henry's law constant data was available for eight benzenoids and a combinatorial descriptor was selected and we obtained the following mono-parametric equation:

$$\log\ H = 2.7611(\pm 0.2724) - 1.9777 \times 10^{-6}\left(\pm 1.7706 \times 10^{-7}\right)\left(\text{A}_\text{s}^- \bar{V}_\text{s}^-\right)^2 \tag{25}$$

Negative coefficient of descriptors demonstrated that if in benzenoids we have more points with negative electrostatic potential with low average, electrostatic attractions between positive and negative regions will be stronger and less molecules can go to gas phase. Fig. SG of the supplementary information shows good agreements between calculated and experimental values.

Bioconcentration factor (log BCF)

Bioconcentration factor is the ratio of a substance's concentration in an organism to its concentration in the ambient water [55]:

$$\text{BCF} = \frac{C_{\text{org}}}{C_\text{w}} \tag{26}$$

where C_{org} is the concentration in target organism (μg/kg) and C_w is the concentration in pure water (μg/l). Bioconcentration factor data was available for 11 benzenoids and after descriptors selection the following equation obtained:

$$\log \text{BCF} = -15.319(\pm 2.2417) + 2.2452(\pm 0.26667)\overline{V}_s^+ \tag{27}$$

Correlation between log BCF and $\overline{V}_\text{s}^+$ demonstrates that there are molecules in organism that have regions with negative electrostatic potential which attract benzenoids molecule into organism. In Eq. (27) phenanthrene is an outlier and the removal of this molecule increase R^2 to 0.926. Fig. SH of the supplementary information presents results for mono-parametric regression.

Photo-induced toxicity (log 1/LT_{50})

This term is used for the phenomenon of increasing the toxicity of certain poly cyclic hydrocarbon such as benzenoids when exposed to UV light due to the formation of the free radicals and subsequent damage of macromolecule and is calculated as:

$$\text{Photo - induced toxicity} = \log\left(\frac{1}{\text{LT}_{50}}\right) \tag{28}$$

where LT_{50} is median lethal time. Anthracene, pyrene, benzo[*a*]pyrene, dibenz[*a*,*h*]anthracene and benzo[*ghi*]perylene are among the most phototoxic compounds whereas phenanthrene and tryphenylene are not phototoxic. Data set included nine benzenoids and after descriptors selection steps, a model with two descriptors created:

$$\begin{aligned}\mathrm{PIT} = 2.9275(\pm 1.3189) - 0.18886(\pm 0.034743)\delta^2_{\mathrm{tot}} \\ + 0.011162(\pm 0.004434)A_{\mathrm{s}}^{+}\end{aligned} \tag{29}$$

Negative coefficient for δ^2_{tot} shows more δ^2_{tot}, decrease photo-induced toxicity and positive coefficient of $\mathrm{A}_{\mathrm{s}}^{+}$ demonstrates that benzenoids with more rings have more photo-induced toxicity because of being $\mathrm{A}_{\mathrm{s}}^{+}$ and $\mathrm{A}_{\mathrm{s}}^{\mathrm{tot}}$ as collinear. As seen in Fig. SI of supplementary information there are moderate agreement between calculated and experimental values for biparametric correlation because of the scarcity data.

Polarographic half-wave reduction potentials ($E_{1/2}$)

Polarographic half-wave reduction potentials data ($E_{1/2}$, in unit of volt) were available for 27 benzenoids, and two descriptors were selected and the following equation was obtained:

$$\begin{aligned}E_{\frac{1}{2}} = 2.9672(\pm 0.79989) - 0.16559(\pm 0.024092)\delta^2_{-} \\ - 0.06761(\pm 0.025846)V_{\mathrm{s,max}}\end{aligned} \tag{30}$$

Larger δ^2_{-} decreases $E_{1/2}$. this is reasonable because presence of points with more negative electrostatic potential repel the electrons. For Eq. (30) benzo[*a*]perylene and dibenzo[*a*,*i*]pyrene are outliers and when they were removed, R^2 increased to 0.822. These descriptors have no collinearity problem ($R^2 = 0.209$). As it's seen in Fig. SJ of supplementary information, there are weak agreement between calculated and experimental values for biparametric correlation because they do not depend on electrostatic interaction only.

Heat of formation (ΔH_{f})

Heat of formation data (in unit of KJ/mol) was available for 20 benzenoids and finally the following model was obtained:

$$(\Delta H_{\mathrm{f}}) = -44.756(\pm 10.618) + 2.3549(0.069023)A_{\mathrm{s}}^{-} \tag{31}$$

Larger benzenoids have larger A_{s}^{-} and this means larger benzenoids have more bonds that results in larger heat of formation. Correlation between A_{s}^{-} and $A_{\mathrm{s}}^{\mathrm{tot}}$ is high ($R^2 = 0.998$) that is a good reason for accuracy of Eq. (31). This correlation is slightly larger than the correlation between A_{s}^{+} and $A_{\mathrm{s}}^{\mathrm{tot}}$ ($R^2 = 0.997$). In Eq. (31) benzo[*a*]pentaphene is an outlier and the removal of this molecule increase R^2 to 0.989. Also Fig. SK of supplementary information shows very good agreements between calculated and experimental values for mono-parametric regression model.

Photolysis half-live (log $t_{1/2}$)

Photolysis is the most important decay process for PAHs. However, it is unlikely to quantify the photochemical transformation for all PAHs because laboratory tests are expensive and time consuming. QSPR models, which correlate the properties of pollutants with their structure descriptors, may be used to study photolysis mechanisms and to efficiently predict photolysis reaction parameters. In this research, we used GIPF descriptors to create a QSPR model for prediction of benzenoids photolysis half-live. Photolysis half-live (in unit of hour) data set were included seven benzenoids and in descriptors selection steps, a combinatorial descriptor was selected that was resulted in the following model:

$$\log t_{1/2} = 3.1673(\pm 0.45628) - 0.00284(\pm 0.000414)\mathrm{A}_{\mathrm{s}}^{+}\bar{V}_{\mathrm{s}}^{+} \tag{32}$$

Equation (32) shows that photolysis half-live decreases for benzenoid hydrocarbons with more positive electrostatic potential points that have high positive electrostatic potential average value. Since large benzenoids have more points with electron cloud deficiency, thus larger benzenoids have less photolysis half-live (Table S1). While few molecules have data, Fig. SL of supplementary information shows good agreements between calculated and experimental values.

Molecular resonance energy (RE)

Resonance energy data was available for 20 benzenoids and after descriptor selection steps, the following equation was obtained:

$$\begin{aligned}\mathrm{RE} = 0.74732(\pm 0.20896) \\ + 9.01 \times 10^{-7}\left(\pm 9.85 \times 10^{-8}\right) \times \left(\mathrm{A}_{\mathrm{s}}^{-}\bar{V}_{\mathrm{s}}^{-}\right)^2\end{aligned} \tag{33}$$

According to this equation, for benzenoids with more negative electrostatic potential area that their average is smaller, resonance energy increases. Benzenoids with more rings have greater A_{s}^{-} and so they have more resonance energy. For Eq. (33) pentacene is an outlier and when we removed it R^2 increased to 0.912. Fig. SM of supplementary information indicates a plot of the cross-calculated versus experimental RE values for all 20 compounds that are studied. From this Fig., it can be also seen that the predicted values are comparatively in poor agreements with the experimental values, as shown by the R^2_{cv} value (only 0.794).

Correlation between RE and log *H*

For both Eqs. (25) and (33), independent variables are the same that this fact demonstrates these two properties are collinear. Five molecules had data for both properties and the following equation was obtained:

$$\mathrm{RE} = 2.0510(\pm 0.0844) - 4259(\pm 0.0806)\log H \qquad (34)$$

$n = 5$, $R^2 = 0.9030$, $R^2_{\mathrm{adj}} = 0.871$, $\mathrm{RMSE} = 0.1595$, $F = 27.9136$, $R^2_{\mathrm{CV}} = 0.7933$, $\mathrm{RMSE_{CV}} = 0.2635$.

Conclusions

The QSPR/QSAR methodologies based on general interaction properties function (GIPF) family descriptors were successfully applied for predicting the physico-chemical properties/biological activities of benzenoid hydrocarbons and these properties/activities depend on the forces that are electrostatic in nature. These compounds can interact through their both positive and negative electrostatic potential regions, up to a similar extent and are lipophilic. Minimum and maximum of R^2_{adj} for QSPR/QSAR models are 0.637 (for $E_{1/2}$) and 0.993 (for log K_{OC}) and F values are between 16.2 (for log 1/LT_{50}) and 1462.1 (for BP). QSPR model for boiling point has the maximum RMSE due to its large boiling points values. $R^2_{\max}$ for five models were larger than 0.4 because a few number of molecules have data for these properties.

References

1. H.J. Lee, J. Villaume, D.C. Cullen, B.C. Kim, M.B. Gu, Biosens Bioelectron **18**, 571 (2003)
2. J.C. Drosos, M. Viola-Rhenals, R. Vivas-Reyes, J Chromatogr A **1217**, 4411 (2010)
3. Q. Jun, S. Chang-Hong, W. Jia, Procedia. Environ Sci **2**, 1429 (2010)
4. S. Tao, X.C. Jiao, S.H. Chen, F.L. Xu, Y.J. Li, F.Z. Liu, Environ Pollut **140**, 13 (2006)
5. J. Beyer, G. Jonsson, C. Porte, M.M. Krahn, F. Ariese, Toxicol Phar **30**, 224 (2010)
6. I. Martorell, G. Perelló, R. Martí-Cid, V. Castell, J.M. Llobet, J.L. Domingo, Environ Int **36**, 242 (2010)
7. R. Ghavami, B. Sepehri, J Chromatogr A **1233**, 116 (2012)
8. F. Liu, Y. Liang, C. Cao, N. Zhou, Anal Chim Acta **594**, 279 (2007)
9. R.J. Hu, H.X. Liu, R.S. Zhang, C.X. Xue, X.J. Yao, M.C. Liu, Z.D. Hu, B.T. Fan, Talanta **68**, 31 (2005)
10. X.J. Yao, A. Panaye, J.P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu, B.T. Fan, J Chem Inf Comput Sci **44**, 1257 (2004)
11. M. Karelson, *Molecular Descriptors in QSAR/QSPR* (Wiley, New York, 2000)
12. R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics, Volumes I & II* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2009)
13. J. Devillers, A.T. Balaban, *Topological Indices and Related Descriptors in QSAR and QSPR* (Eds. Gordon and Breach, Amestrdam, 1999)
14. M.M.C. Ferreira, Chemosphere **44**, 125 (2001)
15. G.N. Lu, Z. Dang, X.O. Tao, C. Yang, X.Y. Yi, Sci Total Environ **373**, 289 (2007)
16. S. Nikolić, A. Miličević, N. Trinajstić, Croat Chem Acta **79**, 155 (2006)
17. P. Politzer, J.S. Murray, Fluid Phase Equilibr **185**, 129 (2001)
18. P. Politzer, J.S. Murray, P. Flodmark, J Phys Chem **100**, 5538 (1996)
19. P. Politzer, J.S. Murray, F.A. Bulat, J Mol. Model **16**, 1731 (2010)
20. P. Jin, T. Brinck, J.S. Murray, P. Politzer, Int J Quantum Chem **95**, 632 (2003)
21. W. Karacher, *Spectral Atlas of Polycyclic Aromatic Compounds*, vol. 2 (Kluwer academic publishers, Dordrecht, 1988), p. 16
22. L.C. Sander, S.A. Wise, Adv Chromatogr **25**, 139 (1986)
23. D. Mackay, W.-Y. Shiu, K.C. Ma, *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate of Organic Compounds*, vol. 2 (Lewis/CRC, Boca Raton, 1992)
24. D. Mackay, D. Calloct, Partitioning and physical properties of PAHs, in *The Handbook of Environmental Chemistry*, vol. 3, Part J. PAHs and related compounds, ed. by A.H. Neilson (Springer, Berlin, 1998), pp. 325–346
25. A.T. Balaban, M. Pompe, J Phys Chem A **111**, 2448 (2007)
26. M. Randić, Chem Rev **103**, 3449 (2003)
27. ISIS Draw 2.3 (MDL Information Systems, Inc., 1990–2000)
28. HyperChem Release 7.1 for Windows Molecular Modeling System Program Package, (HyperCube, 2002)
29. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dappich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. Petersson, P.Y. Aayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rubuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, *Gaussian98, Revision A.5* (Gaussian Inc, Pittsburgh, 1998)
30. H.Y. Xu, J.W. Zou, Q.S. Yu, Y.H. Wang, J.Y. Zhang, H.X. Jin, Chemosphere **66**, 1998 (2007)
31. F.A. Bulat, A. Toro-Labbé, T. Brinck, J.S. Murray, P. Politzer, J Mol Model **16**, 1679 (2010)
32. J.S. Murray, F. Abu-Awwad, P. Politzer, J Phys Chem A **103**, 1853 (1999)
33. Y. Ma, K.C. Gross, C.A. Hollingsworth, P.G. Seybold, J.S. Murray, J Mol Model **10**, 235 (2004)
34. O.G. Gonzalez, J.S. Murray, Z. Peralta-Inga, P. Politzer, Int J Quantum Chem **83**, 115 (2001)
35. P. Kulshrestha, N. Sukumar, J.S. Murray, R.F. Giese, T.D. Wood, J Phys Chem A **113**, 756 (2009)
36. M.N. Hasan, P.C. Jurs, Anal Chem **60**, 978 (1988)
37. J. Olivero, T. Garcia, P. Payares, R. Viva, D. Diaz, E. Daza, P. Geerliger, J Pharm Sci **86**, 625 (1997)
38. R. Ghavami, S. Faham, Chromatographia **72**, 893 (2010)

39. M. Dumarey, A.M.V. Nederkassel, E. Deconinck, Y.V. Heyden, J Chromatogr A **1192**, 81 (2008)
40. J.G. Topliss, R.P. Edwards, J Med Chem **22**, 1238 (1979)
41. K. Varmuza, P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics* (Taylor & Francis Group, LLC, 2009)
42. H. Kubinyi, QSAR: Hansch analysis and related approaches (VCH VerlagsgesellschaftmbH, D-69451 Weinheirn, Federal Republic of Germany, 1993)
43. H. Kubinyi, F.A. Hamprecht, T. Mietzner, J Med Chem **41**, 2553 (1998)
44. D.M. Hawkins, J. Kraker, J Chemometr **24**, 188 (2010)
45. R. Ghavami, A. Najafi, M. Sajadi, F. Djannaty, J Mol Graph Modell **27**, 105 (2008)
46. D.M. Hawkins, S.C. Basak, D. Mills, J Chem Inf Comput Sci **43**, 579 (2003)
47. A. Tropsha, P. Gramatica, V. Gombar, Quant Struct Act Relat Comb Sci **22**, 69 (2003)
48. R. Ghavami, F. Sadeghi, Chromatographia **70**, 851 (2009)
49. C. Hansch, R.P.A. Verma, Eur J Med Chem **44**, 274 (2009)
50. R. Christoph, R. Gerta, M. Markus, J Chem Inf Model **47**, 2345 (2007)
51. J.S. Murray, T. Brinck, P. Politzer, J Phys Chem **97**, 13807 (1993)
52. C. Hansch, J.E. Quinlan, G.L. Lawerence, J. Org. Chem. **33**, 347 (1968)
53. T. Harner, M. Shoeib, J Chem Eng Data **47**, 228 (2002)
54. M. Shoeib, T. Harner, Environ Toxicol Chem **21**, 984 (2002)
55. P. Sang, J.W. Zou, P. Zhou, L. Xu, Chemosphere **83**, 1045 (2011)