

QSAR Applications on Polycyclic Aromatic Hydrocarbons and Some Derivatives

Matías F. Andrada^a, Pablo R. Duchowicz^b and Eduardo A. Castro^b

^aÁrea de Química Física, Departamento de Química, Universidad Nacional de San Luis, Chacabuco 917, 5700 San Luis, Argentina;

^bInstituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina

Abstract: Polycyclic Aromatic Hydrocarbons (PAHs) have been the focus of great attention for a long time owing to their impact on public health and the environment. In this article we overview various QSAR studies on PAHs and some derivatives, that have been developed during the last decade by different authors. Several properties of broad interest are analyzed, such as carcinogenesis, mutagenicity, biocatalytic oxidation, phototoxicity, dermal penetration, biodegradation, and others. The molecular descriptors types and modeling techniques used are briefly commented.

Keywords: QSAR Theory, Polycyclic Aromatic Hydrocarbons, Molecular Descriptors, Multivariable Linear Regression, Cross Validation.

1. INTRODUCTION

Polycyclic Aromatic Hydrocarbons (PAHs) are aromatic multi ring systems produced by an incomplete combustion and pyrolysis of the organic materials [1, 2]. Their occurrence in urban atmosphere largely results of anthropogenic emissions, such as mobile emissions (vehicular, shipping and flying), domestic heating, oil refining, waste incineration, industrial activities, asphalt production, agricultural burning of biomass, etc. [3, 4]. PAH compounds are present in the atmosphere, water, sediments, tobacco smoke, and food [5-7].

PAHs are well-known compounds because of their toxicity, potential to bioaccumulate, and carcinogenic activity [8]. Although these compounds are considered as the main pollutants of air [9], soil acts as the last depository of these chemicals. Their fate in environment includes volatilization, photo and chemical oxidation, adsorption on soil particles, leaching, and microbial degradation [10]. The extent and rate of biodegradation depends on various factors such as the chemical structure of the compound, pH, temperature, oxygen, accessibility of nutrients and the degree of adaptation of the microbial degradation enzyme.

Most of the heaviest PAHs are carcinogenic and have been categorized by the International Agency for Research on Cancer (IARC) as human carcinogens (class 1), probable human carcinogens (class 2A) or possible human carcinogens (class 2B) [11].

Although the PAH structure seems rather homogenous (planar or quasi-planar molecule), the wide variety in the number and position of fused benzene rings and the presence or absence of cyclopentane rings, alkyl substituents, and keto groups gives rise to thousands of different PAHs with various biological activities; many can be identified in environmental samples. In past years, different toxicity studies of PAHs have been carried out by means of the Quantitative Structure-Activity Relationships (QSAR) Theory [12, 13]. The main assumption of QSAR is that the biological activity of a chemical compound is completely determined by its molecular structure. This theory does not offer specific details on

the usually complex mechanism/path of action involved. However, it is possible to get some insight on the underlying mechanism by means of the QSAR-based predicted activities [14, 15].

Within the realms of QSAR, the molecular structure is quantified by using a set of suitable molecular descriptors, which are numbers carrying information on the constitutional, topological, geometrical, hydrophobic, and/or electronic aspects of the chemical structure [16-19]. A set of descriptors can then be statistically correlated to the experimental biological activity, resulting in a model which can be used to find out useful parallelisms. In general, QSAR analyzes are affected by various factors from which the most important are: (a) the composition of the training and test sets; (b) the choice of molecular descriptors that should be representative of the molecular structures and have low colinearity between them; (c) the number of descriptors included in the model; (d) the use of suitable multicriteria modeling methods; and (e) the employment of validation techniques to verify the predictive performance of the developed models [20-24]. The QSAR approach is an important tool for Medicinal Chemistry, and makes drug design more rational by decreasing the number of expensive, time consuming experiments.

In the next section we present various QSAR works on PAHs and some derivatives that have been developed during the last decade by different authors. We describe QSAR studies that report validated models and do not consider published models that are only correlative without validation testing.

2. QSAR STUDIES ON PAHS IN 2003-2012

Zhou *et al.* [25] develop in 2003 a QSAR for exploring the carcinogenesis of PAHs. The Semiempirical Method AM1 implemented in the Mopac 7 package [26] is used to calculate relative thermodynamic stabilities of epoxide and carbonium intermediates of 48 PAHs. All data analysis and regressions are performed using Statistica 8.0 [27] and Molecular Operating Environment (MOE) [28] softwares. Using epoxyl-energy and cationized-energy of two active sites, the Multivariable Linear Regression (MLR) model reasonably predicts the carcinogenic activity of these PAHs with explained variance $R^2 = 0.59$, and shows a good ability to distinguish between carcinogenic and noncarcinogenic PAHs. In this model, two outliers are recognized; if such molecules are excluded from the training set, the new regression model has $R^2 = 0.66$. The

*Address correspondence to this author at the Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina; Tel: ?????????; Fax: ?????????; E-mail: ?????????

variances of Leave One Out Cross Validation (LOO) are: $Q^2 = 0.49$ (whole set of data) and $Q^2 = 0.56$ (excluding two outliers).

Furthermore, the model suggests that double active sites and their distance characteristics are important factors in the chemical carcinogenesis of PAHs. The physical meaning for energies of addition reactions with DNA is also discussed.

In 2004, Niu and Yu [29] establish two QSAR models for the biocatalytic oxidation specific activity of unmodified and chemically modified hemoglobin in the oxidation of different PAHs in 15% acetonitrile. Some fundamental quantum chemical descriptors computed with the PM3 hamiltonian are obtained with Mopac contained in the CS Chem3D Ultra [30]. The Simca software [31] is used to perform a Partial Least Squares (PLS) analysis. The cross-validated Q^2 values for the two optimal QSAR models are 0.78 and 0.75, respectively, indicating a good predictive ability for biocatalytic oxidation specific activity of PAHs. The main factors affecting specific activity of PAHs are most positive net atomic charges on a hydrogen atom (q_{H^+}), largest negative atomic charge on a carbon atom (q_{C^-}), dipole moment (μ), the energy of the highest occupied molecular orbital (ϵ_{homo}), and the square difference between ϵ_{homo} and ϵ_{lumo} ($\Delta\epsilon_{\text{homo-lumo}}^2$), where ϵ_{lumo} is the energy of the lowest unoccupied molecular orbital. The biocatalytic oxidation specific activity of PAHs with high q_{C^-} and $(\epsilon_{\text{homo}} - \epsilon_{\text{lumo}})^2$ values tends to be slow. Increasing q_{H^+} , μ , and ϵ_{homo} values of PAHs leads to increase of specific activity.

A study performed by Ribeiro and Ferreira [32] in 2005 analyzes the photo-induced toxicity in a set of 67 non-substituted PAHs containing 2-7 rings with 5 and 6 carbon atoms. The group of structures used to construct the model (training set) involves only 17 non-substituted PAHs, for which the experimental data for phototoxicity has been determined. Three of these compounds are considered as outliers and excluded from further analysis. The geometry of all PAHs is optimized with the AM1 method of Spartan [33], and used to calculate electronic descriptors: ϵ_{homo} , ϵ_{lumo} , and the gap $\Delta\epsilon_{\text{homo-lumo}} = \epsilon_{\text{homo}} - \epsilon_{\text{lumo}}$. The relationship between such molecular descriptors and the phototoxicity is non-linear, therefore Gaussian type functions are used to linearize it. The model is designed with the aid of the PLS Toolbox of Matlab [34] on mean centered data, and validated with the LOO technique. The best regression equation is:

$$\text{phototoxicity} = -4.3355 + 0.4737 \epsilon_{\text{homo}} + 0.3692 \epsilon_{\text{lumo}} + 1.1081 \Delta\epsilon_{\text{homo-lumo}}$$

The regression analysis results in $R^2 = 0.84$ and $Q^2 = 0.79$. It is applied to predict the phototoxicity of molecules with unknown experimental measurements. Pentaphene, Benzo[b]chrysene and Dibenz[a,j]anthracene are among the potentially phototoxic predicted compounds. A new gap range of 7.2 ± 0.7 eV is proposed for the classification of phototoxic compounds and a larger cutoff of $-\log(ALT) \leq -2.95$ is suggested for the normalized lethal time (ALT). Another study of the same year of Gallegos Saliner and Gironés [35] presents new molecular descriptors for QSAR. It demonstrates that the application of Quantum Similarity Theory together with the classical topological approach in order to obtain Topological Quantum Similarity Indices [36, 37] yields to satisfactory results, which can improve the models obtained by using only similarity measures. This method is applied to predict the dermal penetration of 60 commercially available PAHs, with a structure of 3 up to 7 fused rings. The PLS QSAR model compares fairly well with others from

the literature, involving 5 descriptors with statistical parameters $R^2 = 0.69$ and $Q^2 = 0.65$.

In 2007 Gramatica *et al.* [38] use an homogeneous set of mutagenicity data (TA98 and TA100,+S9) of 32 benzocyclopenta-phenanthrenes and chrysenes, for establishing classification models into mutagenic and non-mutagenic classes. The molecular geometry optimization involves the Molecular Mechanics method of Allinger (MM+) in HyperChem 7.03 [39]. A pool of 1206 molecular descriptors is calculated using the Dragon 5.3 software [40]. Five quantum-chemical descriptors, such as ϵ_{homo} , ϵ_{lumo} , $\Delta\epsilon_{\text{homo-lumo}}$, the ionization potential (IP), and the heat of formation (ΔH_f) are calculated with the Semiempirical PM3 method and added to the pool. The Genetic Algorithm technique provides the best 1-3 Dragon descriptors for modeling the mutagenicity, and the k -Nearest Neighbor (k -NN) and Classification and Regression Tree (CART) methods provide the best classification results. The models found are validated by Cross Validation (Leave-One-Out and Leave-50%-Out) and have good performance, with sensitivity and specificity ranges of 90-100%.

In 2007 we applied the Euler equations for chemical graphs, which are considered an extension of the Euler equations for polyhedra [41]. These equations admit several potential forms of molecular descriptors that can be used in the characterizations of the properties of PAHs in QSAR studies. The nature of these Euler relations is described for hydrocarbon graphs and the descriptors they admit, applying them to predict through MLR 37 boiling points ($R^2_{\text{train}} = 0.98$, $S_{\text{train}} = 12.98$, $Q^2 = 0.98$, $S_{\text{LOO}} = 13.67$), 26 partition coefficients for the n-octanol-water system, K_{ow} ($R^2_{\text{train}} = 0.98$, $S_{\text{train}} = 0.20$, $Q^2 = 0.97$, $S_{\text{LOO}} = 0.22$), and 47 retention time indices for reversed-phase liquid chromatography analysis ($R^2_{\text{train}} = 0.88$, $S_{\text{train}} = 0.33$, $Q^2 = 0.86$, $S_{\text{LOO}} = 0.34$). Final results suggest that these new descriptors can be used to complement others in a QSPR-QSAR study.

In another work of this year, Gramatica *et al.* [42] model a set of mutagenicity data (TA100) for 48 nitro-PAHs, applying OECD principles for QSAR model validation. The proposed MLR models are based on two topological molecular descriptors. Such models are validated for predictivity by both internal and external validation. For the external validation, three different splitting approaches, D-optimal Experimental Design, Self Organizing Maps (SOM) and Random Selection by activity sampling, are applied to the original data set in order to compare these methodologies and to select the best descriptors able to model each prediction set of chemicals independently of the splitting method applied. The applicability domain is verified by the leverage approach.

In 2008 Singh *et al.* [43] apply QSAR studies to describe and predict the mutagenic activity of 48 nitrated PAHs. The molecular structures are drawn using HyperChem [39] and the ACD-Labs software [44]. All the molecular descriptors are computed using Dragon [40] and the Karelson-Chemaxon software [45]. From a larger pool of molecular descriptors (topological indices), a smaller set is found consisting of 3 descriptors. Such a variable selection by made using NCSS software in successive regressions is attempted using maximum- R^2 method. The results are critically discussed using a variety of statistical parameters, revealing that the connectivity and shape type indices together with the distance-based Wiener index (W) play a dominating role in modeling of mutagenicity. The predictive ability of the models is tested with Cross Validation parameters.

Dimitriou-Christidis *et al.* [46] carry out QSAR studies for three Monod-type parameters: the maximal specific biotransformation rate (q_{\max}), the biotransformation affinity coefficient (K_s), and the specific affinity (q_{\max}/K_s). All of these parameters express the kinetics of PAH biotransformation by *Sphingomonas paucimobilis* strain EPA505. The Cerius2 4.9 molecular simulation package [47] is used. The energy of each molecule is minimized using the consistent force field CFF91 [48]. Quantum mechanical descriptors are calculated by Mopac 7 using PM3. The training set contains high-quality experimental values of the kinetic parameters for 20 unsubstituted and methylated PAHs as well as values of 41 meaningful molecular descriptors. Linear regressions are combined with the GA technique of Cerius2, leading to a training set variance around 0.80 for the three kinetic parameters. The Q^2 (25%) values for Leave-25%-Out Cross Validation are: 0.57 (q_{\max}), 0.66 (pK_s) and 0.80 (q_{\max}/K_s). The QSARs are consistent with the hypothesis of membrane transport as being the rate-limiting process in PAH biotransformation by strain EPA505, and a framework for understanding the mechanisms governing biodegradation at the molecular level.

In another project of 2008, Wang *et al.* [49] use Gene Expression Programming (GEP) as a novel machine learning technique. The GEP is used to build a nonlinear QSAR model for the prediction of the Percent of Applied Dose Dermal Absorbed (PADA) over 24 hours for PAHs. This model includes three descriptors in the MLR, selected from the descriptors pool by an Heuristic Method (HM). The GEP method produces a nonlinear quantitative model with quality $R^2_{\text{train}} = 0.85$ and $RMSE_{\text{train}} = 4.70$, and $R^2_{\text{test}} = 0.83$ and $RMSE_{\text{test}} = 7.65$. $RMSE$ is the root mean square error. It is shown that the GEP predicted results are in good agreement with experimental ones.

Papa *et al.* [50] use the experimental data of a mutagenicity test on human B-lymphoblastoid cells (alternative to the Ames bacterial test) for a set of 70 oxo-, nitro- and unsubstituted PAHs, detected in particulate matter, for developing QSAR classification models. The methods used are k -NN, k -Nearest Neighbour and CART, based on different theoretical molecular descriptors calculated with Dragon 5.3 and chosen by GA. The best models are validated for predictivity both externally and internally. For external validation, Self Organizing Maps (SOM) are applied to split the original data set. The best models, developed on the training set alone, show good predictive performance also on the prediction set chemicals (sensitivity 69.2-87.1%, specificity 62.5-87.5%). The classification of PAHs according to their mutagenicity, based only on a few theoretical molecular descriptors, allows a preliminary assessment of the human health risk, and the prioritization of these compounds.

In a work published in 2009, Toropov *et al.* [51] use optimal descriptors calculated with SMILES [44] for the QSAR analysis of the mutagenic potency (TA100) of nitrated polycyclic aromatic hydrocarbons. The blocking of rare SMILES attributes, by means of an appropriate selection of the *LimN* parameter, is able to improve the predictive potential of the model, i.e. improve the statistical characteristics for the external test set. The statistical characteristics found is: $N_{\text{train}} = 28$, $R^2_{\text{train}} = 0.90$, $Q^2 = 0.89$, $S_{\text{train}} = 0.55$ and $N_{\text{test}} = 20$, $R^2_{\text{test}} = 0.85$, $S_{\text{test}} = 0.70$. However, the *LimN* should be properly chosen: zero value of the *LimN* can lead to overtraining, i.e. an excellent model for the training set, but a poor one for the test set; and viceversa: if the *LimN* is too large, then the model can be poor for the training and test sets.

Another QSAR of 2009 of Wang *et al.* [52] involves the photoinduced toxicity (EC_{50}) of 19 PAHs on two aquatic species. PLS regression of the Simca-S package [31] and molecular descriptors calculated with DFT in Gaussian 03 [53] are employed for model development. Two models are established and their high R^2_{train} (0.82 and 0.92) and Q^2 (0.74 and 0.86) values indicate their acceptable goodness-of-fit, robustness and internal predictive power. The average molecular polarizability (α), $\Delta\epsilon_{\text{homo-lumo}}$, lowest triplet excitation energy (E_{T1}) and vertical electron affinity at the lowest excited triplet (VEA_{T1}) are the main molecular structural factors. The following QSAR models are developed using DFT-based descriptors:

Model 1:

$$\log_{10} EC_{50} = -9.721 \cdot 10^{-3} \alpha + 9.017 \cdot 10^{-2} \Delta\epsilon_{\text{homo-lumo}}^2 + 1.331 \cdot 10^{-1} E_{T1} - 1.849 VEA_{T1} + 9.976$$

Model 2:

$$\log_{10} EC_{50} = -6.396 \cdot 10^{-3} \alpha + 9.463 \cdot 10^{-2} \Delta\epsilon_{\text{homo-lumo}}^2 - 1.087 VEA_{T1} - 1.118 \cdot 10^{-1} E_{T1}^2 + 4.443 \cdot 10^{-1} \epsilon_{\text{homo}} + 5.819$$

Polarizability, which determines the partition of PAHs between water and lipid, governs the photoinduced toxicity of selected PAHs. Moreover, the photoinduced toxicity increases with the decrease of $\Delta\epsilon_{\text{homo-lumo}}$ probably due to a better spectral overlap. The parameter VEA_{T1} , that characterizes the ability of PAH anion radical (PAH⁻) generation from excited triplet state PAH, is also related with the photoinduced toxicity. This investigation will make us gain more insight into the photoinduced toxicity mechanism and assess the applicability of various DFT-based descriptors to toxicological QSAR.

Immunoassays are regarded as a possible alternative or supplement for measuring PAHs in the environment. Since there are too many potential cross-reactants for PAHs immunoassays, it is difficult to determine all the cross-reactivities (CRs) by experimental tests. In 2010, Zhang *et al.* [54] investigate the relationship between CR and the physicochemical properties of 40 PAHs and related compounds, using the CR data from a commercial enzyme-linked immunosorbent assay (ELISA) kit test. Two techniques are applied for predicting the CR of PAHs in this ELISA kit, MLR and Comparative Molecular Field Analysis (CoMFA). Parabolic regression indicates that the CRs are significantly correlated with pK_{ow} ($R^2_{\text{train}} = 0.64$, $N_{\text{train}} = 23$, $P < 0.0001$), suggesting that hydrophobic interactions play an important role in the antigen-antibody binding and the cross-reactions in this ELISA test. The CoMFA procedures are performed with a Silicon Graphics workstation using Sybyl 6.91 [55]. The CoMFA model obtained shows that the CRs of the PAHs are correlated with the 3D structure of the molecules ($R^2_{\text{train}} = 0.87$, $N_{\text{train}} = 37$, $Q^2 = 0.66$, $P < 0.0001$). The contributions of the steric and electrostatic fields to CR are 40.4% and 59.6%, respectively. Both QSAR models satisfactorily predict the CR in this PAH immunoassay kit, and help in understanding the mechanisms of antigen-antibody interaction.

Kobeticova *et al.* [56] publish an article in 2011 that compares the toxicity of selected two- and three-ringed PAHs (naphthalene, phenanthrene, and anthracene) and their N-heterocyclic analogs containing one (quinoline, acridine, and phenanthridine) and two (quinoxaline, phenazine, and 1,10-phenanthroline) nitrogen atoms on the survival and reproduction of *E. crypticus* in artificial soil.

When toxicity is based on molar concentrations in pore water ($\mu\text{mol.L}^{-1}$), it significantly increases with increasing K_{ow} value. This relationship indicates nonpolar narcosis as the general toxicity mechanism of the tested compounds. The concentration at which 50% of adult survival (LC_{50}) and 50% inhibition of the reproductive output (EC_{50}) is observed, is calculated by logistic regression analysis with the Least Squares method [57]. The relationships between the properties of compounds and their toxicities are analyzed by the Pearson correlation, while all the statistical analyses are done using Statistica 8.0 software [27]. The multivariate statistical method, PLS implemented in Simca-p+12 [58], is employed to identify additional relationships between physicochemical descriptors and the toxicity data besides K_{ow} . The QSAR models found reveal that the ionization potential (IP) of the tested PACs is as important for the explanation of their toxic effects as their pK_{ow} . For the group of compounds used in this study (i.e. 3 PAHs and 6 NPAHs), this parameter seems rational to be included in QSAR models, because for NPAHs, dissociation (dependent on their ionization potential) affects not only their behavior in soil but also their toxicity. The QSAR models are successfully validated by permutation test and no outliers from these models are found. The QSAR models for 9 tested PAHs lead to the following statistical parameters: $R^2_{train} = 0.80$, $Q^2 = 0.67$ for $\log_{10} LC_{50} = -0.53 \log_{10} K_{ow} + 1.52 IP - 10.76$ and $R^2_{train} = 0.81$, $Q^2 = 0.72$ for $\log_{10} EC_{50} = -0.55 \log_{10} K_{ow} + 1.58 IP - 11.53$.

A study of 2011 by Li *et al.* [59] performs an integrated Molecular Docking and QSAR approach in order to further investigate the binding interactions (K_b) between hydroxylated PAHs (HO-PAHs) and calf thymus DNA (CT-DNA). The binding mode for the HO-PAHs to CT-DNA is analyzed with CDocker, which is incorporated into Discovery Studio 2.5 through the Dock Ligands protocol [60]. Molecular descriptors are chosen using Dragon 2.1 [18] and Gaussian 09 packages [53], which describe hydrophobic, electronic and steric properties of molecules. All the initial geometries of the compounds are optimized with PM3, then optimized at the hybrid HF-DFT B3LYP/6-31G(d,p) level. Solvent (water) effects are taken into consideration implicitly, including the integral equation formulation of the PCM model (IEFPCM) [61]. The frequency analysis is performed on the optimized geometries to ensure that the systems have no imaginary vibration frequencies. Molecular size, polarizability and electrostatic potential are important factors for the binding interactions between HO-PAHs and DNA. The HO-PAHs molecules with smaller molecular size and higher electrostatic interaction tend to have larger binding constants with DNA. The optimal QSAR model found is:

$$\log_{10} K_b = 5.77 - 2.76 SIC2 - 1.43 MATS5p + 2.62 \cdot 10^{-1} Mor29e - 1.45 E2s + 5.36 I$$

The statistical quality of the model is: $N_{train} = 17$, $R^2_{train} = 0.75$, $Q^2 = 0.72$, $N_{test} = 7$, $R^2_{test} = 0.66$, $S_{test} = 0.11$. Therefore, the developed QSAR model has good robustness, predictive ability and mechanistic interpretability, which could be applied to predict the binding constants of other HO-PAHs.

In another work, Li *et al.* [62] establish a QSAR model for the depuration rate constants (k_d) of 28 PAHs, 8 Polybrominated Diphenyl Ethers (PBDEs) and 28 Polychlorinated Biphenyls (PCBs) in mussels (*Elliptio complanata*). All the quantum chemical descriptors are computed by Gaussian 09. Initial geometries of the compounds are optimized with PM3, then optimized at the hybrid

HF-DFT B3LYP/6-31G(d, p) level. The model is obtained with PLS regression in Simca-S, following the guidelines for development and validation of QSAR models. The best regression model is:

$$\log_{10} k_d = 1.34 \cdot 10^{-1} - 1.07 \cdot 10^{-3} \omega - 4.86 \cdot 10^{-3} \alpha - 7.65 \bar{V}_s - 4.31 \cdot 10^{-2} \tau$$

The statistical quality achieved is: $N_{train} = 51$, $R^2_{train} = 0.95$, $Q^2 = 0.95$, $N_{test} = 13$, $R^2_{test} = 0.89$, $S_{test} = 0.16$. The electrophilicity index, molecular polarizability, the averages of the negative potentials on the molecular surface and the balance parameter of surface potential are key parameters governing the pK_d values in the QSAR model, which indicate that pK_d is mainly related to the partition ability, electrostatic interactions, and van der Waals interactions of compounds. The developed QSAR model shows good robustness, predictive ability and mechanistic interpretability, which could be potentially applied to predict the depuration rates of other PAHs, PBDEs and PCBs.

In 2012, the same group [63] integrates Molecular Docking with QSAR in order to understand the relationship between molecular structural features and estrogenic activity, and to predict the binding affinity of PAHs to the estrogen receptor α ($ER\alpha$). Molecular descriptors that describe hydrophobic, electronic and steric properties of molecules are chosen for describing the interaction between PAHs and $ER\alpha$, calculated using Dragon 2.1 and Gaussian 09 softwares. The initial geometries of the compounds are optimized with PM3, then reoptimized at the hybrid HF-DFT B3LYP/6-31G(d,p) level. Solvent (water) effects are taken into consideration implicitly, including the IEFPCM model. Five descriptors ($E1s$, $MATS1v$, $L3s$, $Mor12v$, and $RDF020e$) are included in the QSAR, which indicate that the estrogenic activity is related to molecular size, van der Waals volumes, shape profiles, polarizabilities and electropological states. The optimal QSAR model is:

$$\log_{10} ER\alpha = -5.18 + 6.90 E1s + 1.91 \cdot 10^1 MATS1v - 6.53 L3s - 8.97 \cdot 10^{-1} Mor12v - 4.35 \cdot 10^{-2} RDF020e$$

The statistical quality achieved is: $N_{train} = 29$, $R^2_{train} = 0.94$, $S_{train} = 0.19$, $Q^2 = 0.85$, $N_{test} = 7$, $R^2_{test} = 0.58$, $S_{test} = 0.90$. Comparatively, the developed QSAR model has good robustness, predictive ability and mechanistic interpretability.

A QSAR study of Xu *et al.* [64] includes the biodegradation activity ($pt_{1/2}$) of 17 PAHs. The quantum chemical calculations of equilibrium geometries and vibration frequency are carried out with Gaussian 03 at the B3LYP/6-31+G(d,p) level of theory. The QSAR equation is established by using SPSS statistical analysis software package [65]. The main descriptors influencing $pt_{1/2}$ values that are found with the Stepwise Multiple Linear Regression (SMLR) method are: the wagging vibration frequency of the whole molecule ($Freq$), the number of benzene rings (nB), ϵ_{lumo} , the energy of the next homo (ϵ_{nhomo}) and the biggest infrared frequency intensity ($IRIn$). The equation that has better predictive ability is:

$$pt_{1/2} = 12.049 - 0.013 Freq - 0.159 nB - 5.329 \epsilon_{lumo} + 16.528 \epsilon_{nhomo} - 0.003 IRIn$$

The statistical quality obtained is: $R^2_{train} = 0.90$, $RMSE_{train} = 0.041$, $Q^2 = 0.91$, $RMSE_{LOO} = 0.039$. This QSAR model shows that the biodegradation activity is closely related to the molecular structure: the chemical bond strength of benzene ring plays an important role in biodegradation process, and low molecu-

lar weight PAHs are more degradable than compounds having higher molecular weights.

In another study of the same year, Xu and Li [66] establish a relationship between the chemical structures and biodegradation rates (k_b) of 22 PAHs that range in size from two to four rings, including compounds containing 5-carbon rings and compounds with alkyl substituents. Because the biodegradation reactions in this study happened in an aqueous system degraded by *Sphingomonas yanoikuyae*, solvent effects are considered using IEFPCM. The geometries of all molecules are optimized using DFT at the B3LYP/6-31+G(d,p) level in Gaussian 03. The QSAR equation is established by using SPSS statistical analysis software package. The molecular descriptors chosen with SMLR include: the in-plane bending vibration frequency of the conjugated ring of PAHs ($Freqp$), the largest negative atomic charges on a carbon atom (Q_C^-), and the amount of six-carbon benzenoid rings $N1$:

$$k_b = -0.653 + 0.001 \text{ } Freqp + 0.068 \text{ } Q_C^- + 0.049 \text{ } N1$$

The successfully developed QSAR model has the following statistics: $R^2_{train} = 0.87$, $RMSE_{train} = 0.001$, $Q^2(9\%) = 0.89$, $RMSE_{L90} = 0.0012$. The fact that a bending frequency is more important than the homo or lumo energies in predicting k_b suggests that the bending of benzene ring might play an important role in the enzymatic catalysis of the initial oxidation step.

Xu *et al.* [67] study the relationship between the soot-water partition coefficient ($\log_{10} K_{SC}$) and the structure of persistent organic pollutants (POPs), in order to analyze the sorption behavior of such compounds. The initial geometries are optimized with Mopac 6.0 implemented in the Vega package using the AM1 method [68], and reoptimized at the HF/6-31G* level with Gaussian 98. The correlation is established with SMLR in SPSS 10.0. The results show that for Diesel soot (SRM 1650) the best descriptors are the sum of the surface minima values of the electrostatic potential together with the molecular surface area. The following equation is established for the calibration set:

$$\log_{10} K_{SC} = 8.996 + 2.811 \text{ } 10^{-2} \text{ } A_s + 26.91 \text{ } \epsilon_{homo} + 4.729 \text{ } \bar{V}_s^- + 1.486 \text{ } V_{s,max}$$

The quality of the derived QSAR is: $N_{train} = 17$, $R^2_{train} = 0.98$, $S_{train} = 0.16$, $Q^2 = 0.96$.

Another project of 2012 of Al-Fahemi [69] deals with the prediction of the photoinduced toxicity ($\log_{10} EC_{50}$) of PAH compounds toward two aquatic species, *Daphnia magna* and *Scenedesmus vacuolatus*. The geometries of the PAHs are fully optimized and calculated with B3LYP and the 3-21G basis set. Various quantum chemical descriptors such as ϵ_{homo} , ϵ_{lumo} , $\Delta\epsilon_{homo-lumo}$, electronegativity, chemical potential, chemical hardness, softness index, electrophilicity and polarizability are used along with physico-

chemical descriptors to construct useful QSAR models. The SMLR method is employed to generate linear models, and it is carried out with SPSS. The descriptors of the photoinduced toxicity in *Daphnia magna* are: ϵ_{lumo} and molar refractivity, leading to $N_{train} = 14$, $R^2_{train} = 0.82$, $S_{train} = 0.57$, $Q^2 = 0.67$ ($\log_{10} EC_{50} = 7.690 - 0.0528 \text{ } MR + 1.368 \text{ } \epsilon_{lumo}$). For the case of *Scenedesmus vacuolatus*, the best descriptors are $\Delta\epsilon_{homo-lumo}$, molar refractivity and electronegativity, with quality $N_{train} = 12$, $R^2_{train} = 0.92$, $S_{train} = 0.31$, $Q^2 = 0.80$, $Q^2 = 0.67$ ($\log_{10} EC_{50} = 7.303 - 0.029 \text{ } MR - 2.213 \text{ } \epsilon_{lumo}$). According to the author, 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.

3. CONCLUSION

In this review article we present various QSAR studies on PAHs and some derivatives that are established in the period 2003-2012, in order to reveal the state of art of this subject. The analyzed properties have broad interest and involve carcinogenesis, mutagenicity, biocatalytic oxidation, phototoxicity, dermal penetration, biodegradation, and others. The molecular descriptors types and modeling techniques used are briefly commented.

(Table 1) summarizes the reported QSAR models, including the number of compounds used for training and testing external predictive capability together with their statistical quality. It is appreciated that the number of compounds used during past years for establishing the structure-activity relationships has been small for most cases. In addition, the descriptors used in such studies have not been exhaustively chosen from a pool containing thousands of them, which would lead to a better result. Therefore, in line with these observations, among the perspectives for future QSAR research on PAHs, we encourage new studies to address the following main points: i. to increase the number of PAH compounds for obtaining more representative models; ii. to use external test sets of compounds for an appropriate validation of such models; iii. to search for the best descriptors by exploring a pool that contains thousands of them; iv. to employ alternative modeling techniques, such as nonlinear methods of Artificial Neural Networks and Support Vector Machines; v. to use the developed QSAR for *a priori* prediction of the analyzed properties for compounds with unknown observations; vi. to expand the use of QSAR models to other physicochemical and biological properties of interest. Finally, the most important result would be to apply the developed models to predict compounds that have no determined experimental values, and then to experimentally verify if the predictions achieved are correct or not, in other words, we call for an interdisciplinary communication between modelers and experimentalists. We expect that current review may serve as a guide for new QSAR projects on PAHs.

Table 1. Details of QSAR Models Established on PAHs

Authors	Sets	Statistical Parameters
Zhou <i>et al.</i> [25]	$N_{train} = 48$ structures	$R^2_{train} = 0.59$; $Q^2 = 0.49$ (whole dataset) $R^2_{train} = 0.66$; $Q^2 = 0.56$ (excluding 2 outliers)
Niu and Yu [29]	$N_{train} = 10$ structures	$R^2_{train} = 0.82$; $Q^2 = 0.78$ (model 1) $R^2_{train} = 0.79$; $Q^2 = 0.75$ (model 2)

Table 1. contd....

Authors	Sets	Statistical Parameters
Ribeiro and Ferreira [32]	$N_{train} = 17$ structures	$R^2_{train} = 0.84$; $Q^2 = 0.79$
Gallegos Saliner and Gironés [35]	$N_{train} = 60$ structures	$R^2_{train} = 0.69$; $Q^2 = 0.65$
Gramatica <i>et al.</i> [38]	$N_{train} = 32$ structures	90-100% range of sensitivity and specificity
Duchowicz <i>et al.</i> [41]	$N_{train1} = 37$ structures	$R^2_{train} = 0.98$; $Q^2 = 0.98$ (model 1)
	$N_{train2} = 26$ structures	$R^2_{train} = 0.98$; $Q^2 = 0.97$ (model 2)
	$N_{train3} = 47$ structures	$R^2_{train} = 0.88$; $Q^2 = 0.86$ (model 3)
Gramatica <i>et al.</i> [42]	$N_{train} = 48$ structures	$R^2_{train} = 0.90$; $Q^2 = 0.87$
Singh <i>et al.</i> [43]	$N_{train} = 48$ structures	$R^2_{train} = 0.92$; $Q^2 = 0.90$
Dimitriou-Christidis <i>et al.</i> [46]	$N_{train} = 20$ structures	$R^2_{train} = 0.84$; $Q^2 = 0.57$ (for q_{max})
		$R^2_{train} = 0.82$; $Q^2 = 0.66$ (for pK_s)
		$R^2_{train} = 0.88$; $Q^2 = 0.80$ (for q_{max}/pK_s)
Wang <i>et al.</i> [49]	$N_{train} = 20$ structures	$R^2_{train} = 0.85$; $R^2_{test} = 0.83$
	$N_{test} = 10$ structures	
Wang <i>et al.</i> [52]	$N_{train1} = 12$ structures	$R^2_{train} = 0.82$; $Q^2 = 0.74$ (model 1)
	$N_{train2} = 14$ structures	$R^2_{train} = 0.92$; $Q^2 = 0.86$ (model 2)
Zhang <i>et al.</i> [54]	$N_{train1} = 23$ structures	$R^2_{train} = 0.64$ (model 1); $Q^2 = 0.60$
	$N_{train2} = 37$ structures	$R^2_{train} = 0.87$ (model 2); $Q^2 = 0.83$
Kobeticova <i>et al.</i> [56]	$N_{train} = 9$ structures	$R^2_{train} = 0.80$; $Q^2 = 0.67$ (for LC_{50})
		$R^2_{train} = 0.81$; $Q^2 = 0.72$ (for EC_{50})
Li <i>et al.</i> [59]	$N_{train} = 17$ structures	$R^2_{train} = 0.75$; $R^2_{test} = 0.66$
	$N_{test} = 7$ structures	
Li <i>et al.</i> [62]	$N_{train} = 51$ structures	$R^2_{train} = 0.95$; $R^2_{test} = 0.89$
	$N_{test} = 13$ structures	
Li <i>et al.</i> [63]	$N_{train} = 29$ structures	$R^2_{train} = 0.94$; $R^2_{test} = 0.58$
	$N_{test} = 7$ structures	
Xu <i>et al.</i> [64]	$N_{train} = 17$ structures	$R^2_{train} = 0.91$; $Q^2 = 0.90$
Xu and Li [66]	$N_{train} = 22$ structures	$R^2_{train} = 0.87$; $Q^2(9\%) = 0.89$
Xu <i>et al.</i> [67]	$N_{train} = 17$ structures	$R^2_{train} = 0.98$; $Q^2 = 0.96$
	$N_{train} = 8$ structures	
Al-Fahemi [69]	$N_{train1} = 14$ structures	$R^2_{train} = 0.82$; $Q^2 = 0.67$ (model 1)
	$N_{train2} = 12$ structures	$R^2_{train} = 0.92$; $Q^2 = 0.80$ (model 2)

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank the financial support provided by the National Research Council of Argentina (CONICET) PIP11220100100151 project and to Ministerio de Ciencia, Tecnología e Innovación Productiva for the electronic library facilities.

ABBREVIATIONS

ANN	=	Artificial Neural Networks
CART	=	Classification and Regression Tree
CoMFA	=	Comparative Molecular Field Analysis
DFT	=	Density Functional Theory
\mathcal{E}_{homo}	=	Energy of the highest occupied molecular orbital
\mathcal{E}_{lumo}	=	Energy of the lowest unoccupied molecular orbital

$\Delta\epsilon_{\text{homo-lumo}}$	=	Difference between ϵ_{homo} and ϵ_{lumo}
GA	=	Genetic Algorithms
GEP	=	Gene Expression Programming
HF	=	Hartree-Fock
HO-PAHs	=	Hidroxylyated- Polycyclic Aromatic Hydrocarbons
IEFPCM	=	Integral Equation Formulation of the Polarized Continuum Model
K_{ow}	=	Partition coefficient for the n-octanol-water system
LOO	=	Leave One Out Cross Validation
LNO	=	Leave N% Out Cross Validation
MLR	=	Multivariable Linear Regression
N	=	Number of molecules
PAHs	=	Polycyclic Aromatic Hydrocarbons
PBDEs	=	Polybrominated Diphenyl Ethers
PCBs	=	Polychlorinated Biphenyls
PLS	=	Partial Least Squares
Q^2	=	Explained variance of LOO
$Q^2(n\%)$	=	Explained variance of LNO
QSAR	=	Quantitative Structure-Activity Relationship
QSPR	=	Quantitative Structure-Property Relationship
R	=	Correlation coefficient of the model
R^2	=	Explained variance
RMSE	=	Root Mean Square Error
S	=	Standard deviation of the model
SMLR	=	Stepwise Multiple Linear Regression
test	=	Test set
train	=	Training set

REFERENCES

- [1] Lubov, O.; Mikhail, S.; Grigoriy, T.; Yury, Z. On-line solid-phase extraction and high performance liquid chromatography determination of polycyclic aromatic hydrocarbons in water using polytetrafluoroethylene capillary. *Talanta*, **2007**, *4*, 1386-1391.
- [2] Manahan, S.E. *Environmental Chemistry*. 9th ed. CRC Press: New York, **2009**.
- [3] Marchand, N.; Besombes, J.L.; Chevron, N.; Masclet, P.; Aymoz, G.; Jaffrez, J.L. Polycyclic aromatic hydrocarbons (PAHs) in the atmospheres of two French alpine valleys: sources and temporal patterns. *Atmos. Chem. Phys.*, **2004**, *4*, 1167-1181.
- [4] Ravindra, K.; Sokhi, R.; Van Grieken, R. Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. *Atmos. Environ.*, **2008**, *42*, 2895-2921.
- [5] Freeman, D.J.; Cattell, F.C.R. Wood burning as a source of atmospheric polycyclic aromatic hydrocarbons. *Environ. Sci. Technol.*, **1990**, *24*, 1581-1585.
- [6] Lijinsky, W. The formation and occurrence of polynuclear aromatic hydrocarbons associated with food. *Mut. Res.*, **1991**, *259*, 251-261.
- [7] Nishioka, M.; Chang, H.C.; Lee, V. Structural characteristics of polycyclic aromatic hydrocarbon isomers in coal tars and combustion products. *Environ. Sci. Technol.*, **1986**, *20*, 1023-1027.
- [8] Mastral, A.M.; Callen, M.S. A review on polycyclic aromatic hydrocarbons emission from energy generation. *Environ. Sci. Technol.*, **2000**, *34*, 3051-3057.
- [9] Haritash, A.K.; Kaushik, C.P. Seasonal and spatial occurrence and distribution of respirable particulate-bound atmospheric polycyclic aromatic hydrocarbons in Hisar city and their potential health-risks. *Asian J. Wat. Environ. Pollut.*, **2011**, *8*, 73-80.
- [10] Wild, S.R.; Jones, K.C. Polynuclear aromatic hydrocarbons in the United Kingdom environment: a preliminary source inventory and budget. *Environ. Pollut.*, **1995**, *88*, 91-108.
- [11] IARC (International Agency for Research on Cancer), 2012. Agents Classified by the IARC Monographs. Available: In: List of Classifications by Alphabetical Order, vol. 1e105. <http://monographs.iarc.fr/ENG/Classification/ClassificationsAlphaOrder.pdf> [accessed January 2012].
- [12] Mekenyan, O.G.; Ankley, G.T.; Veith, G.D.; Call, D.J. QSARs for photoinduced toxicity: 1. Acute lethality of polycyclic aromatic hydrocarbons to *Daphnia magna*. *Chemosphere*, **1995**, *28*, 567-582.
- [13] Veith, G.D.; Mekenyan, O.G.; T., A.G.; Call, D.J. A QSAR Analysis of substituent effects on the photoinduced acute toxicity of PAHs. *Chemosphere*, **1995**, *30*, 2129-2142.
- [14] Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*. Wiley-Interscience: New York, **2008**.
- [15] Puzyn, T.; Leszczynski, J.; T., C.M. *Recent Advances in QSAR Studies: Methods and Applications*. First Edition ed. Springer: New York, **2009**.
- [16] Diudea, M.V.E. *QSPR/QSAR Studies by Molecular Descriptors*. Nova Science Publishers: New York, **2001**.
- [17] Katritzky, A.R.; Lobanov, V.S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.*, **1995**, *24*, 279-287.
- [18] Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*. Wiley-VCH: Weinheim, **2009**.
- [19] Trinajstić, N. *Chemical Graph Theory*. CRC Press: Boca Raton (FL), **1992**.
- [20] Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graphics Modell.*, **2002**, *20*, 269-276.
- [21] Goodarzi, M.; Duchowicz, P. R.; Wu, C. H.; Fernández, F. M.; Castro, E. A. New Hybrid Genetic Based Support Vector Regression as QSAR Approach for Analyzing Flavonoids-GABA(A) Complexes. *J. Chem. Inf. Model.*, **2009**, *49*, 1475-1485.
- [22] Hawkins, D.M.; Basak, S. C.; Mills, D. Assessing Model Fit by Cross Validation. *J. Chem. Inf. Model.*, **2003**, *43*, 579-586.
- [23] Putz, M.V.; Putz, A.-M.; Lazea, M.; Ienciu, L.; Chiriac, A. Quantum-SAR Extension of the spectral-SAR algorithm. Application to polyphenolic anticancer bioactivity. *Int. J. Mol. Sci.*, **2009**, *10*, 1193-1214.
- [24] Putz, M.V.; Ionaşcu, C.; Putz, A.M.; Ostafe, V. Alert-QSAR. Implications for Electrophilic Theory of Chemical Carcinogenesis. *Int. J. Mol. Sci.*, **2011**, *12*, 5098-5134.
- [25] Zhou, Z.; Dai, Q.; Gu, T. A QSAR Model of PAHs Carcinogenesis Based on Thermodynamic Stabilities of Biactive Sites. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 615-621.
- [26] *Mopac 7, Stewart Computational Chemistry*, www.openmopac.net.
- [27] *Statistica 8.0 for Windows*, <http://www.statsoft.com/#>.
- [28] *Molecular Operating Environment (MOE)*, www.chemcomp.com: 2003.
- [29] Niu, J.; Yu, G. Molecular structural characteristics governing biocatalytic oxidation of PAHs with haemoglobin. *Environ. Tox. Pharm.*, **2004**, *18*, 39-45.
- [30] *CS Chem3D Ultra version 6.0*, <http://cs-chem3d-ultra.software.informer.com>.
- [31] Umetri, A.B.; Erisoft, A.B. *Simca-S. Version 6.0*, 2006.
- [32] Ribeiro, F.A.L.; Ferreira, M.M.C. QSAR model of the phototoxicity of polycyclic aromatic hydrocarbons. *J. Mol. Struct. THEOCHEM*, **2005**, *719*, 191-200.
- [33] *Spartan, version 5.1.3*, Wavefunction Inc., 1995-2001.
- [34] *Matlab 5.0 for Windows*, The MathWorks, Inc.: 1995.
- [35] Gallegos Saliner, A.; Gironés, X. Topological quantum similarity measures: applications in QSAR. *J. Mol. Struct. THEOCHEM*, **2005**, *727*, 97-106.
- [36] Carbó-Dorca, R.; Robert, D.; Amat, L.I.; Jirones, X.; Besalú, E. *Molecular Quantum Similarity in QSAR and Drug Research*. Springer-Verlag: Berlin, **2000**.
- [37] Lobato, M.; Amat, L.; Besalú, E.; Carbó-Dorca, R. Structure-Activity Relationships of a Steroid Family using Quantum Similarity Measures and Topological Quantum Similarity Indices. *Quant. Struct.-Act. Relat.*, **1997**, *16*, 465-472.
- [38] Gramatica, P.; Papa, E.; Marrocchi, A.; Minuti, L.; Taticchi, A. Quantitative structure-activity relationship modeling of polycyclic aromatic hydrocarbon mutagenicity by classification methods based on holistic theoretical molecular descriptors. *Ecotox. Environ. Safe.*, **2007**, *66*, 353-361.
- [39] *HyperChem 7.03 for Windows*, Hypercube, Inc. <http://www.hyper.com>: Gainesville, 2002.
- [40] *Dragon 5.3 for Windows*, Taletta Srl, Milan, (Italy).
- [41] Duchowicz, P.R.; Bucknum, M.J.; Castro, E.A. New molecular descriptors based upon the Euler equations for chemical graphs. *J. Math. Chem.*, **2007**, *41*, 193-208.
- [42] Gramatica, P.; Pilutti, P.; Papa, E. Approaches for externally validated QSAR modelling of nitrated polycyclic aromatic hydrocarbon mutagenicity. *SAR and QSAR Environ. Res.*, **2007**, *18*, 169-178.
- [43] Singh, J.; Singh, S.; Shaik, B.; Deeb, O.; Sohani, N.; Agrawal, V.K.; Khadikar, P.V. Mutagenicity of Nitrated Polycyclic Aromatic Hydrocarbons: A QSAR Investigation. *Chem. Biol. Drug Des.*, **2008**, *71*, 230-243.
- [44] *ACD-Labs software for calculating the referred physicochemical parameters; ChemSketch 3.0*, <http://www.acdlabs.com>.
- [45] Karelson, M. *Chemaxon softwares for the calculation of topological indices*, <http://www.chemaxon.com>.
- [46] Dimitriou-Christidis, P.; Autenrieth, R.L.; Abraham, M.H. Quantitative Structure-Activity Relationships For Kinetic Parameters of Polycyclic Aromatic Hydrocarbon Biotransformation. *Environ. Toxicol. Chem.*, **2008**, *27*, 1496-1504.
- [47] *Cerius2*, Accelrys Inc., <http://accelrys.com>.

- [48] Maple, J.R.; Hwang, M.J.; Stockfisch, T.P.; Dinur, U.; Waldman, M.; Ewig, C.S.; Hagler, A.T. Derivation of class II force fields: 1. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.*, **1994**, *15*, 162-182.
- [49] Wang, T.; Si, H.; Chen, P.; Zhang, K.; Yao, X. QSAR models for the dermal penetration of polycyclic aromatic hydrocarbons based on gene expression programming. *QSAR Comb. Sci.*, **2008**, *27*, 913-921.
- [50] Papa, E.; Pilutti, P.; Gramatica, P. Prediction of PAH mutagenicity in human cells by QSAR classification. *SAR and QSAR Environ. Res.*, **2008**, *19*, 115-127.
- [51] Toropov, A.A.; Toropova, A.P.; Benfenati, E. Simplified Molecular Input Line Entry System-Based Optimal Descriptors: Quantitative Structure-Activity Relationship Modeling Mutagenicity of Nitrated Polycyclic Aromatic Hydrocarbons. *Chem. Biol. Drug Des.*, **2009**, *73*, 515-525.
- [52] Wang, Y.; Chen, J.; Li, F.; Qin, H.; Qiao, X.; Hao, C. Modeling photoinduced toxicity of PAHs based on DFT-calculated descriptors. *Chemosphere*, **2009**, *76*, 999-1005.
- [53] Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Montgomery, J.A.; Vreven, T.; Kudin, K.N.; Burant, J.C.; Millam, J.M.; Iyengar, S.S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, R.J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, v.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A. *Gaussian 09, Gaussian Inc., Pittsburgh, PA*, 2009.
- [54] Zhang, Y.; Ma, Y.; Gao, Z.; Dai, S. Predicting the cross-reactivities of polycyclic aromatic hydrocarbons in ELISA by regression analysis and CoMFA methods. *Anal. Bioanal. Chem.*, **2010**, *397*, 2551-2557.
- [55] Sybyl 6.91, Tripos, <http://www.tripos.com/index.php?family=modules,SimplePage...&page=SYBYL-X>.
- [56] Kobeticová, K.; Simek, Z.; Brezovský, J.; Hofman, J. Toxic effects of nine polycyclic aromatic compounds on *Enchytraeus crypticus* in artificial soil in relation to their properties. *Ecotox. Environ. Safe.*, **2011**, *74*, 1727-1733.
- [57] Sverdrup, L.E.; Kelley, A.E.; Krogh, P.H.; Nielsen, T.; Jensen, J.; Scott-Fordsmand, J.J.; Stenersen, J. Effect of eight polycyclic aromatic compounds on the survival and reproduction of the springtail *Folsomia finetaria* (Collembola, Isotomidae). *Environ. Toxicol. Chem.*, **2001**, *20*, 1332-1338.
- [58] *SIMCA-P+12*, Umetrics AB, Sweden. <http://www.umetrics.com/>
- [59] Li, F.; Li, X.; Liu, X.; Zhang, L.; You, L.; Zhao, J.; Wu, H. Noncovalent interactions between hydroxylated polycyclic aromatic hydrocarbon and DNA: Molecular docking and QSAR study. *Environ. Tox. Pharm.*, **2011**, *32*, 373-381.
- [60] *CDocker*, incorporated into Discovery Studio 2.5. <http://accelrys.com/>.
- [61] Huetz, P.; Kamarulzaman, E.E.; Wahab, H.A.; Mavri, J. Chemical reactivity as a tool to study carcinogenicity: reaction between estradiol and estrone 3,4-quinones ultimate carcinogens and guanine. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 310-314.
- [62] Li, F.; Liu, X.; Zhang, L.; You, L.; Wu, H.; Li, X.; Zhao, J.; Yu, J. QSAR studies on the depuration rates of polycyclic aromatic hydrocarbons, polybrominated diphenyl ethers and polychlorinated biphenyls in mussels (*Elliptio complanata*). *SAR and QSAR Environ. Res.*, **2011**, *5-6*, 561-573.
- [63] Li, F.; Wua, H.; Li, L.; Li, X.; Zhao, J.; Peijnenburg, W.J.G.M. Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. *Ecotox. Environ. Safe.*, **2012**, *80*, 273-279.
- [64] Xu, X.; Li, X.G.; Sun, S.W. A QSAR study on the biodegradation activity of PAHs in aged contaminated sediments. *Chemom. Intel. Lab. Syst.*, **2012**, *114*, 50-55.
- [65] *SPSS for Windows*, standard version, SPSS Inc., <http://www-01.ibm.com/software/analytics/spss>.
- [66] Xu, X.; Li, X.G. QSAR for Predicting Biodegradation Rates of Polycyclic Aromatic Hydrocarbons in Aqueous Systems. *Chinese J. Struct. Chem.*, **2012**, *31*, 1212-1221.
- [67] Xu, H.Y.; Zou, J.W.; Min, J.Q.; Wang, W. A quantitative structure-property relationship analysis of soot-water partition coefficients for persistent organic pollutants. *Ecotox. Environ. Safe.*, **2012**, *80*, 1-5.
- [68] Pedretti, A.; Villa, L.; Vistoli, G. Vega: a versatile program to convert, handle and visualize molecular structure on windows-based PCs. *J. Mol. Graphics Model.*, **2002**, *21*, 47-49.
- [69] Al-Fahemi, J.H. The use of quantum-chemical descriptors for predicting the photoinduced toxicity of PAHs. *J. Mol. Model.*, **2012**, *18*, 4121-4129.