ELSEVIER

Journal of Molecular Structure (Theochem) 663 (2003) 109–126

THEO CHEM

www.elsevier.com/locate/theochem

# QSPR models of boiling point, octanol–water partition coefficient and retention time index of polycyclic aromatic hydrocarbons

Fabiana Alves de Lima Ribeiro, Márcia Miguel Castro Ferreira*

*Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP 13083-970, Brazil*

Received 17 February 2003; accepted 15 August 2003

**Abstract**

A Quantitative Structure–Property Relationship (QSPR) analysis and study of polycyclic aromatic hydrocarbons (PAHs) is presented. Three physicochemical properties related to their environmental impact are studied: boiling point (bp), octanol–water partition coefficient (log $K_{ow}$) and retention time index (RI) for reversed-phase liquid chromatography analysis. The geometry of all PAHs were optimized by the semi-empirical method AM1 and used to calculate thermodynamic, electronic, steric and topological descriptors: HOMO and LUMO energies and the GAP between them, molecular hardness, polarizability, atomic charges, connectivity index, volume and surface area among others. After variable selection, principal component regression (PCR) and partial least squares (PLS) with leave-one-out crossvalidation were used for building the regression models.

The regression coefficients obtained for the models were 0.995 (PCR and PLS) for bp, 0.975 (PCR) and 0.976 (PLS) for log $K_{ow}$, and 0.898 (PCR and PLS) for RI. Finally, the models were used to predict these properties for those compounds for which experimental measurements are still unknown.


*Keywords:* Polycyclic aromatic hydrocarbons; Principal component regression; Partial least squares; Molecular descriptors; AM1; WHIM-3D; Physicochemical properties

## 1. Introduction

There are currently more than 50 million known chemical compounds and the chemical industry produces tons of synthetic compounds annually. Since the industrial revolution, increasing amount of chemical products and waste materials of all kinds are being released into the environment either during production, storage, transport or disposal. Once in the environment, they frequently interfere with naturally occurring chemical reactions and cycles. Some of them have received enormous attention in the last decades due to their toxicity, bioaccumulation and persistence in the environment [1]. Their existence promotes important changes in the planet's life, originating a process with two totally opposite consequences. On one hand allowing an exceptional human standard of life, but on the other hand leading to several environmental degradation processes that can result in serious consequences for the ecosystem

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

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

0166-1280/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2003.08.107

and, ultimately, for life on the planet. To understand and evaluate the environmental impact of chemicals, it is necessary to have information about the chemical, biochemical and photochemical processes that generate or consume these products, their transport, adsorption, persistence and metabolism within living organisms. This context emerges the field of environmental chemistry, that encompasses their chemical reactions, and how they can contaminate the air, water, soil and sediment, and the treatment of residues generated from the innumerable industrial processes used to sustain the actual human standard of life. The effect of these chemicals in the environment depends mainly upon two factors [1]: (a) environmental conditions such as temperature, flux and accumulation of air, water and solid matter, and sediment composition; (b) the physicochemical properties of pollutants, which will influence the way these compounds disperse or accumulate in the environment. Due to the complexity of these factors, the present level of knowledge in this area is not very high and there is an essential need for study and research to gain a deeper understanding of the effects of pollutants and their chemistry.

A powerful tool to help in this task is chemometrics, which uses statistical and mathematical methods to extract maximum of information from a data set. Its usefulness relies on the fact that the complex interactive processes in the environment are per se multidimensional. Quantitative structure–activity and structure–property relationships (QSAR/QSPR) use chemometric methods to describe how a given biological activity or a physicochemical property varies as a function of molecular descriptors describing the chemical structure of the molecule. Thus, it is possible to replace costly biological tests or experiments of a given physicochemical property (especially when involving hazardous and toxically risky materials or unstable compounds) with calculated descriptors, which can in turn be used to predict the responses of interest for new compounds [2]. Chemometrics has provided new insight into the philosophy and theory behind QSAR/QSPR modeling. It has been used to estimate properties such as density [3], boiling point (bp) [4,5], solubility, octanol–water partition coefficient, Henry's law constant [6–8], vapor pressure [9] and toxicity of chemicals [7,9–13], for a series of analogue compounds such as polychlorinated biphenyls, dibenzofuranes, chlorobenzenes, polychlorinated dioxins, polycyclic aromatic hydrocarbons (PAHs), fatty acids, etc.

The aim of this work is to obtain QSPR of three physicochemical properties: bp, octanol–water partition coefficient (log $K_{ow}$ or log $P$), and retention time index (RI) for reversed-phase liquid chromatography (LC) analysis, for a set of 67 non-substituted fused PAHs, a special class of chemicals that has been of concern to the scientific community due to their pollutant potential. PAHs are found in both urban and rural areas, due to the burning of wood and coal, exhaust of gasoline and diesel from combustion engines, the smoking of tobacco, and other combustion processes in which the carbon fuel is not completely converted to CO or $CO_2$. They are produced by saturated hydrocarbons under oxygen deficient conditions, by degradation or incomplete combustion of organic materials. Under these conditions, hydrogen consumption is favored, and the exceeding carbon atoms will be arranged in the most thermodynamically favorable way consisting of condensed aromatic rings [14].

The production of PAHs generates a variety of compounds with similar structures and properties. One of the analytical methods for isomer separation is reversed-phase LC on stationary phases chemically modified with octadecilsilane. This technique has been successfully used to resolve complex mixtures of PAHs [15]. The authors utilized the log $I$ (RI) to study the selectivity of LC. This index is calculated by Eq. (1), where $x$ refers to the solute, $n$ and $n+1$ the minor and major standards of elution, and $R$ represent the corrected elution volume. For the index calculation, the following standard values are assumed: benzene = 1; naphthalene = 2; phenanthrene = 3; benz[*a*]anthracene = 4; benzo[*b*]chrysene = 5; dibenzo[*b.def*]chrysene = 6

$$\log I_x = \log I_n + \frac{\log R_x - \log R_n}{\log R_{(n+1)} - \log R_n} \quad (1)$$

PAHs are well known for their mutagenic and carcinogenic characteristics. In general, a larger number of condensed rings usually exhibit greater environmental significance because most of them are carcinogenic and extremely resistant to enzymatic degradation [14,16].

On natural bodies of water, PAHs can contaminate water as well as sediment, depending on their hydrophobicity. The most hydrophobic PAHs have the tendency to adhere to sediment, or biological tissues from aquatic living organisms, while those not so hydrophobic have the tendency to solubilize in an aqueous phase [1,6]. The hydrophobicity is expressed by the octanol–water partition coefficient ($K_{ow}$), which estimates the solubility in both aqueous and organic phases (in general *n*-octanol is used) according to Eq. (2)

$$K_{ow} = \frac{C_{org}}{C_{aq}} \tag{2}$$

Since values of $K_{ow}$ may vary by several orders of magnitude, it is usually expressed in the logarithmic form [2].

The $\log K_{ow}$ is essential for understanding the transport mechanisms and distribution of compounds into the environment, for example, the mechanism that involves drug absorption by transport through a biological membrane, or the process involving the deposition of a pollutant into bodies of water. The contamination of soil by PAHs takes place mainly by atmospheric deposition and by sewage water or sewage sludge, and the transport of these compounds occurs by diffusion. Consequently, the soil permanence is very much related to water solubility of each compound [17,18].

Human contamination takes place mainly by inhalation of contaminated particles, cutaneous absorption, or orally through contaminated food. The chemicals adhered to particles are partly dissolved in the lung and metabolized there. The metabolites can act on cells, or can be taken into the blood and liver, or can be excreted unchanged. The majority of these compounds tend to bioaccumulate, and only a small fraction of them is eliminated by urine, either in the metabolized or original form [14,19].

## 2. Methodology

*Data set*. In this work, a set of 67 non-substituted PAHs containing from 2 to 7 fused rings with five and six carbon atoms were studied. Their chemical structures are listed in Fig. 1.

Experimental values of bp were taken from the work of Karcher et al. [20], for the octanol–water partition coefficient they were taken from the handbook by Mackay et al. [1] and for RI from the work of Sander and Wise [15]. They are listed in Table 1, together with the values predicted by the principal component regression (PCR) and partial least squares (PLS) models.

*Descriptors calculations*. Firstly, the geometry of all molecules was optimized and the electronic and steric descriptors were calculated using the semi-empirical method AM1 [21] implemented in the Spartan software [22]. These descriptors are: volume, surface area and molecular ovality [23,24], atomic charges, dipole moment, frontier orbitals energies—HOMO and LUMO energies and their GAP, molecular electronegativity, molecular hardness and polarizability [25–28].

The topological descriptors [29,30] were calculated by the WHIM-3D program [31]. These descriptors contain information about the whole molecule in terms of size, shape, symmetry and atom distribution. These indices are calculated from the cartesian $(x, y, z)$ coordinates of a molecule within different weighting schemes in a straightforward manner.

*Variable selection*. The data matrix $X(n \times m)$ with $n = 67$ rows and $m = 168$ columns corresponds, respectively, to the number of molecules investigated and molecular descriptors calculated. This diversity of molecular descriptors was evaluated in order to find those that provide the best regression model for bp, $\log K_{ow}$ and RI. The physicochemical properties are, in general, intrinsically multidimensional. Therefore, the use of just one molecular descriptor would not be sufficient to supply all the necessary information to describe the data set. Some descriptors give valuable information about the influence of electronic, others about geometrical or topological, and others about hydrophobic features upon bp, $\log K_{ow}$ and RI. In this work, the molecular descriptors were selected in such a way that they represent the features necessary to quantify these physicochemical properties. The correlation of all 168 descriptors with a given molecular property was calculated and those with small or no correlation (smaller than a given cut-off) to any of the physicochemical properties were discarded. Among those highly correlated to themselves, the ones that can be most easily interpreted were selected. The data

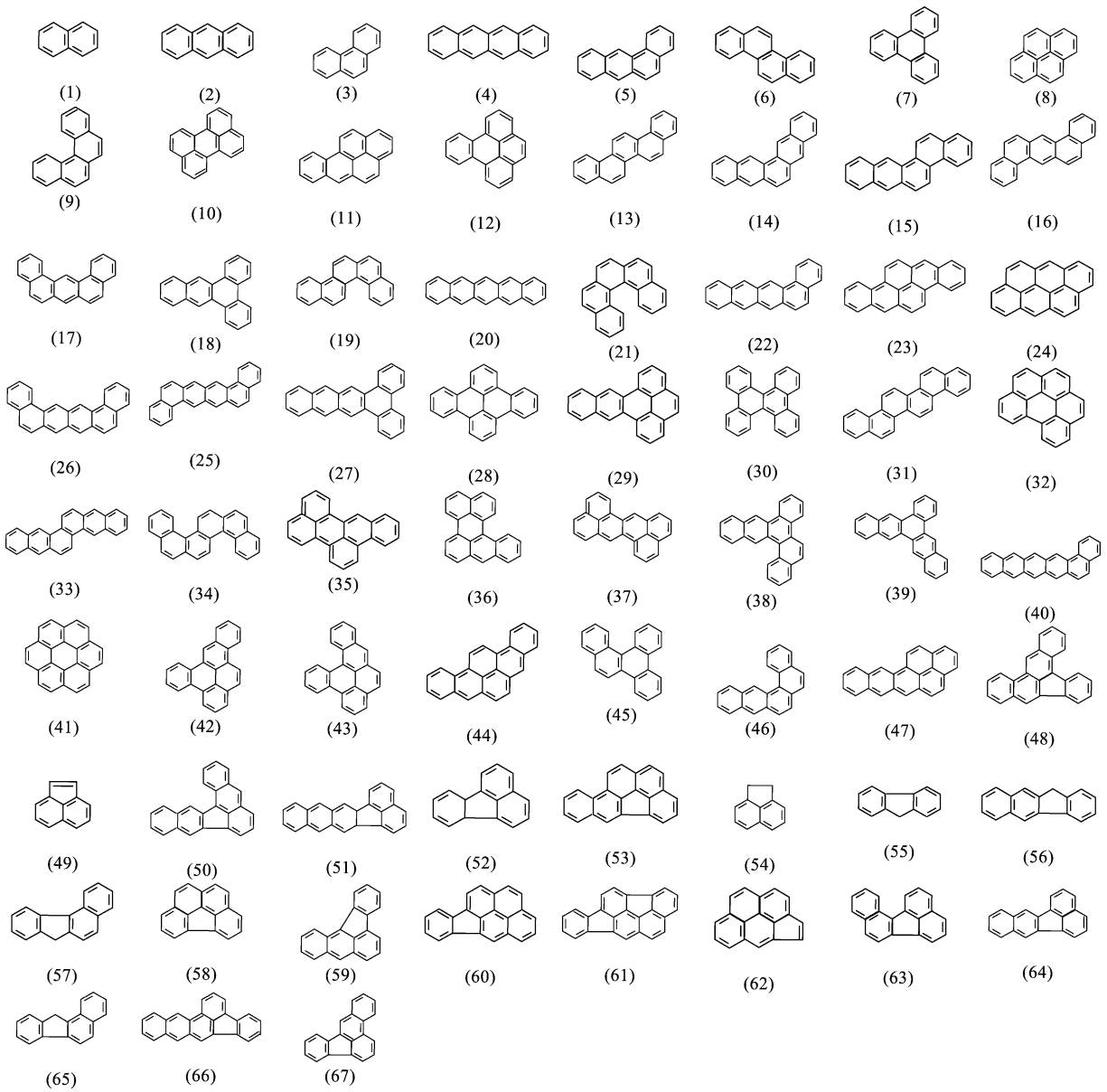


Fig. 1. Chemical structures of PAHs used in this work.

set was reduced to 10 descriptors which are listed in Table 2, and their respective correlations with the properties of interest can be observed in Fig. 2.

*Modeling and prediction*. QSPR models for bp, log $K_{ow}$ and RI were constructed by PCR and PLS methods [32–34] on autoscaled data and validated by leave-one-out crossvalidation. The best model was selected to predict the properties for some PAHs for which the experimental values still are unknown. The statistical parameters used to assess the quality of the models are the Prediction Error Sum of Squares (PRESS) of validation (Eq. (3)), the Standard Error of Validation (SEV) (Eq. (4)) and finally the standard and crossvalidated

Table 1
Experimental and predicted values for bp, log $K_{ow}$ and RI

| CAS name | bp | | | log $K_{ow}$ | | | RI | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 Naphthalene | 218.00 | | | 3.37 | | | | 2.28 | 2.28 |
| 2 Anthracene | 340.00 | | | 4.54 | | | 3.2 | | |
| 3 Phenanthrene | 338.00 | | | 4.57 | | | 3 | | |
| 4 Naphthacene | 440.00 | | | 5.76 | | | 4.51 | | |
| 5 Benz[*a* ]anthracene | 435.00 | | | 5.91 | | | 4 | | |
| 6 Chrysene | 431.00 | | | 5.86 | | | 4.1 | | |
| 7 Triphenylene | 429.00 | | | 5.49 | | | 3.7 | | |
| 8 Pyrene | 393.00 | | | 5.18 | | | 3.58 | | |
| 9 Benzo[*c* ]phenanthrene | | 430.79 | 431.23 | | 5.71 | 5.70 | 3.64 | | |
| 10 Perylene | 497.00 | | | 6.25 | | | 4.33 | | |
| 11 Benzo[*a* ]pyrene | 496.00 | | | 6.04 | | | 4.53 | | |
| 12 Benzo[*e* ]pyrene | 493.00 | | | | 6.04 | 6.05 | 4.28 | | |
| 13 Picene | 519.00 | | | | 6.77 | 6.74 | | 4.51 | 4.51 |
| 14 Pentaphene | | 534.13 | 534.38 | | 6.77 | 6.76 | 4.67 | | |
| 15 Benzo[*b* ]chrysene | | 534.73 | 534.96 | | 6.77 | 6.75 | 5 | | |
| 16 Dibenz[*a,h* ]anthracene | 535.00 | | | 6.75 | | | 4.73 | | |
| 17 Dibenz[*a,j* ]anthracene | 531.00 | | | | 6.77 | 6.75 | 4.56 | | |
| 18 Benzo[*b* ]triphenylene | 535.00 | | | | 6.76 | 6.74 | 4.4 | | |
| 19 Benzo[*c* ]chrysene | | 535.25 | 535.46 | | 6.77 | 6.74 | 4.45 | | |
| 20 Pentacene | | 533.54 | 533.81 | | 6.80 | 6.77 | | 5.40 | 5.40 |
| 21 Dibenzo[*c.g* ]phenanthrene | | 535.54 | 535.73 | | 6.76 | 6.74 | 4.07 | | |
| 22 Benzo[*a* ]naphthacene | | 534.12 | 534.36 | | 6.77 | 6.76 | 4.99 | | |
| 23 Dibenzo[*b.def* ]chrysene | 596.00 | | | | >6.75 | >6.75 | 6 | | |
| 24 Dibenzo[*def.mno* ]chrysene | 547.00 | | | | >6.75 | 6.53 | 5.08 | | |
| 25 Dibenzo[*a.j* ]naphthacene | | 638.97 | 638.96 | | >6.75 | >6.75 | | 5.77 | 5.76 |
| 26 Dibenzo[*a.l* ]naphthacene | | 638.98 | 638.97 | | >6.75 | >6.75 | | 5.75 | 5.75 |
| 27 Dibenzo[*a.c* ]naphthacene | | 639.84 | 639.80 | | >6.75 | >6.75 | | 5.71 | 5.71 |
| 28 Dibenzo[*el* ]naphthacene | | 590.38 | 590.67 | | >6.75 | >6.75 | | 4.84 | 4.84 |
| 29 Dibenzo[*de.gr* ]naphthacene | | 589.02 | 589.35 | | >6.75 | >6.75 | 4.92 | | |
| 30 Dibenzo[*g.p* ]chrysene | | 640.90 | 640.82 | | >6.75 | >6.75 | | 5.36 | 5.36 |
| 31 Benzo[*c* ]picene | | 640.18 | 640.13 | | >6.75 | >6.75 | | 5.44 | 5.43 |
| 32 Benzo[*ghi* ]perylene | 542.00 | | | 6.50 | | | 4.76 | | |
| 33 Dibenzo[*b.k* ]chrysene | | 639.12 | 639.10 | | >6.75 | >6.75 | | 5.80 | 5.80 |
| 34 Dibenzo[*cl* ]chrysene | | 640.11 | 640.06 | | >6.75 | >6.75 | | 5.34 | 5.34 |
| 35 Benzo[*b* ]perylene | | 589.57 | 589.88 | | >6.75 | >6.75 | 5.04 | | |
| 36 Benzo[*a* ]perylene | | 589.23 | 589.55 | | >6.75 | >6.75 | | 5.54 | 5.54 |
| 37 Dibenzo[*de.mn* ]naphthacene | | 588.76 | 589.10 | | >6.75 | >6.75 | | 5.65 | 5.65 |
| 38 Naphtho[2.3-*g* ]chrysene | | 641.43 | 641.33 | | >6.75 | >6.75 | | 5.48 | 5.48 |
| 39 Benzo[*h* ]pentaphene | | 639.85 | 639.80 | | >6.75 | >6.75 | | 5.34 | 5.33 |
| 40 Benzo[*a* ]pentacene | | 638.49 | 638.50 | | >6.75 | >6.75 | | 6.13 | 6.12 |
| 41 Coronene | 590.00 | | | 6.75 | | | | 4.91 | 4.92 |
| 42 Naphtho[1.2.3.4-*def* ]chrysene | 592.00 | | | | >6.75 | >6.75 | 4.97 | | |
| 43 Dibenzo[*def.p* ]chrysene | 595.00 | | | | >6.75 | >6.75 | 4.89 | | |
| 44 Benzo[*rst* ]pentaphene | 594.00 | | | | >6.75 | >6.75 | 5.73 | | |
| 45 Benzo[*g* ]chrysene | | 535.78 | 535.97 | | 6.76 | 6.74 | 4.27 | | |
| 46 2.3:5.6-Dibenzophenanthrene | | 534.71 | 534.93 | | 6.78 | 6.75 | 4.33 | | |
| 47 Naphtho[2.1.8-*qra* ]naphthacene | | 588.41 | 588.76 | | >6.75 | >6.75 | 5.87 | | |
| 48 Dibenz[*a.e* ]aceantrylene | | 587.31 | 587.70 | | >6.75 | >6.75 | 4.9 | | |
| 49 Acenaphthylene | 270.00 | | | 4.00 | | | | 2.86 | 2.85 |

(*continued on next page*)

Table 1 (*continued*)

| CAS name | bp | | | log $K_{ow}$ | | | RI | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 50 Dibenzo[*a.k*]fluoranthene | | 587.03 | 587.43 | | >6.75 | >6.75 | 4.9 | | |
| 51 Naphtho[2.3-*k*]fluoranthene | | 586.55 | 586.96 | | >6.75 | >6.75 | | 5.11 | 5.11 |
| 52 Fluoranthene | 383.00 | | | 5.22 | | | 3.37 | | |
| 53 Dibenzo[*k.mno*]fluoranthene | | 539.48 | 539.31 | | 6.54 | 6.56 | | 4.53 | 4.53 |
| 54 1.2-Dihydroacenaphthylene | 279.00 | | | 3.92 | | | | 2.57 | 2.56 |
| 55 9H–fluorene | 294.00 | | | 4.18 | | | 2.7 | | |
| 56 Benzo[*b*]fluorene | 398.00 | | | 5.75 | | | 3.84 | | |
| 57 Benzo[*c*]fluorene | 406.00 | | | | 5.40 | 5.39 | 3.49 | | |
| 58 Benzo[*ghi*]fluoranthene | 422.00 | | | | 5.49 | 5.51 | | 3.78 | 3.78 |
| 59 Benzo[*a*]aceanthrylene | | 485.78 | 485.48 | | 6.08 | 6.08 | 4.22 | | |
| 60 Indeno[1.2.3.*cd*]pyrene | 534.00 | | | | 6.54 | 6.55 | 4.84 | | |
| 61 Indeno[1.2.3-*cd*] fluoranthene | 531.00 | | | | 6.57 | 6.58 | 4.93 | | |
| 62 Cyclopenta[*cd*]pyrene | 439.00 | | | | 5.40 | 5.43 | | 4.12 | 4.13 |
| 63 Benzo[*j*]fluoranthene | 480.00 | | | | 6.07 | 6.07 | 4.24 | | |
| 64 Benzo[*k*]fluoranthene | 481.00 | | | 6.00 | | | 4.42 | | |
| 65 Benzo[*a*]fluorene | 403.00 | | | 5.40 | | | 3.72 | | |
| 66 Dibenz[*e.k*]acephenanthrylene | | 641.67 | 641.79 | | >6.75 | >6.75 | 5.27 | | |
| 67 Benzo[*b*]fluoranthene | 481.00 | | | 5.80 | | | 4.29 | | |

(1) Bp experimental values, extracted from Ref. [20]; (2) bp predicted values by PLS model with two factors; (3) Bp predicted values by PCR model with three factors; (4) log $K_{ow}$ experimental values, extracted from Ref. [1]; (5) log $K_{ow}$ predicted values by PLS model with two factors; (6) log $K_{ow}$ predicted values by PCR model with two factors; (7) RI experimental values, extracted from Ref. [15]; (8) RI predicted values by PLS model with two factors; (9) RI predicted values by PCR model with two factors.

correlation coefficients $R^2$ and $Q^2$, given by Eqs. (5) and (6), respectively, [32,34–36]

$$\mathrm{PRESS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

$$\mathrm{SEV} = \sqrt{\frac{\mathrm{PRESS}}{n}} \tag{4}$$

$$R^2 = 1 - \left( \frac{\mathrm{PRESS}_{\mathrm{cal}}}{\sum_{i=1}^{n} ((y_i - \bar{y})^2)} \right) \tag{5}$$

$$Q^2 = 1 - \left( \frac{\mathrm{PRESS}}{\sum_{i=1}^{n} ((y_i - \bar{y})^2)} \right) \tag{6}$$

In these equations, $n$ is the number of compounds used for crossvalidation, $y_i$ is the experimental value of the physicochemical property for the $i$th sample and $\hat{y}_i$ is the value predicted by the model built without sample $i$. $\mathrm{PRESS}_{\mathrm{cal}}$ is the prediction error sum of squares for all samples included in the model. The models were built on autoscaled data and the classical QSPR regression equation can be obtained by the use of the scaled regression coefficients, mean and standard deviation of each original descriptor. The unscaling algorithm is presented in Appendix A. Multivariate data analysis was carried out by the software Pirouette 2.02 [37] and the PLS Toolbox [38] for MATLAB [39].

## 3. Results and discussion

### 3.1. Boiling point (bp)

The experimental data of 36 compounds listed in Table 1 were used to construct the regression models. From 10 molecular descriptors, three were chosen to model bp: volume ($V$), molecular weight (MW) and Randic connectivity index ($R$) [40,41]. The regression analysis on autoscaled data resulted in a correlation

Table 2
Molecular descriptors selected to construct the QSPR models

| CAS name | $E_{LUMO}$ | $E_{HOMO}$ | GAP | Area | Volume | $\eta$ | $\alpha$ | MW | $R$ | Wiener |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 Naphthalene | −0.2650 | −8.7099 | 8.4449 | 159.0527 | 131.9503 | 4.2225 | 14.5141 | 128.18 | 6.812 | 109 |
| 2 Anthracene | −0.8417 | −8.1212 | 7.2795 | 206.4759 | 177.2617 | 3.6397 | 21.2966 | 178.24 | 9.38 | 279 |
| 3 Phenanthrene | −0.4085 | −8.6171 | 8.2086 | 203.1945 | 176.7936 | 4.1043 | 18.7282 | 178.24 | 9.38 | 271 |
| 4 Naphthacene | −1.2321 | −7.7488 | 6.5167 | 253.8928 | 222.5818 | 3.2583 | 27.3665 | 228.3 | 11.949 | 569 |
| 5 Benz[*a* ]anthracene | −0.8116 | −8.2079 | 7.3963 | 250.6318 | 222.1094 | 3.6982 | 24.5932 | 228.3 | 11.949 | 553 |
| 6 Chrysene | −0.6762 | −8.3697 | 7.6935 | 246.9786 | 221.5272 | 3.8468 | 23.7076 | 228.3 | 11.949 | 545 |
| 7 Triphenylene | −0.4532 | −8.6584 | 8.2052 | 243.0391 | 220.8467 | 4.1026 | 22.3106 | 228.3 | 11.949 | 513 |
| 8 Pyrene | −0.9225 | −8.0692 | 7.1467 | 217.8428 | 193.5795 | 3.5733 | 23.0164 | 202.26 | 10.535 | 362 |
| 9 Benzo[*c* ]phenanthrene | −0.6456 | −8.4438 | 7.7982 | 245.6475 | 221.1480 | 3.8991 | 23.3918 | 228.3 | 11.949 | 529 |
| 10 Perylene | −1.1508 | −7.8598 | 6.7090 | 258.7597 | 238.2124 | 3.3545 | 28.0041 | 252.32 | 13.104 | 654 |
| 11 Benzo[*a* ]pyrene | −1.1142 | −7.9173 | 6.8031 | 262.2384 | 238.6522 | 3.4015 | 27.7382 | 252.32 | 13.105 | 680 |
| 12 Benzo[*e* ]pyrene | −0.8580 | −8.2149 | 7.3569 | 258.5826 | 238.0639 | 3.6784 | 26.0020 | 252.32 | 13.106 | 652 |
| 13 Picene | −0.7209 | −8.3487 | 7.6278 | 290.8768 | 266.2941 | 3.8139 | 27.5208 | 278.36 | 14.518 | 963 |
| 14 Pentaphene | −0.8400 | −8.2022 | 7.3622 | 298.0454 | 267.3996 | 3.6811 | 28.3664 | 278.36 | 14.519 | 979 |
| 15 Benzo[*b* ]chrysene | −0.9948 | −8.0511 | 7.0563 | 294.3653 | 266.8390 | 3.5281 | 29.2342 | 252.32 | 13.104 | 971 |
| 16 Dibenz[*a.h* ]anthracene | −0.8041 | −8.2570 | 7.4529 | 294.6701 | 266.9324 | 3.7264 | 28.0668 | 278.36 | 14.518 | 971 |
| 17 Dibenz[*a,j* ]anthracene | −0.8736 | −8.1916 | 7.3180 | 294.3490 | 266.7377 | 3.6590 | 28.4418 | 278.36 | 14.519 | 955 |
| 18 Benzo[*b* ]triphenylene | −0.8319 | −8.2255 | 7.3936 | 290.0115 | 266.0922 | 3.6968 | 28.1692 | 278.36 | 14.520 | 907 |
| 19 Benzo[*c* ]chrysene | −0.6931 | −8.3898 | 7.6967 | 291.2930 | 266.3563 | 3.8483 | 27.3357 | 278.36 | 14.521 | 931 |
| 20 Pentacene | −1.5500 | −7.4414 | 5.8914 | 301.2299 | 267.9507 | 2.9457 | 33.2215 | 278.36 | 14.522 | 1011 |
| 21 Dibenzo[*c.g* ]phenanthrene | −0.6732 | −8.3498 | 7.6766 | 290.0853 | 266.0904 | 3.8383 | 27.3692 | 278.36 | 14.523 | 899 |
| 22 Benzo[*a* ]naphthacene | −1.1857 | −7.8407 | 6.6550 | 298.0321 | 267.4128 | 3.3275 | 30.5484 | 278.36 | 14.524 | 987 |
| 23 Dibenzo[*b.def* ]chrysene | −1.3630 | −7.6784 | 6.3154 | 306.1395 | 283.5631 | 3.1577 | 32.9926 | 302.38 | 15.673 | 1142 |
| 24 Dibenzo[*def.mno* ]chrysene | −1.4067 | −7.6315 | 6.2248 | 277.1642 | 255.5712 | 3.1124 | 31.0343 | 276.34 | 14.259 | 839 |
| 25 Dibenzo[*a.j* ]naphthacene | −1.1349 | −7.9321 | 6.7972 | 342.2179 | 312.2648 | 3.3986 | 33.7286 | 328.42 | 17.087 | 1581 |
| 26 Dibenzo[*a.l* ]naphthacene | −1.1352 | −7.9345 | 6.7993 | 342.1814 | 312.2526 | 3.3997 | 33.7210 | 328.42 | 17.087 | 1565 |
| 27 Dibenzo[*a.c* ]naphthacene | −1.1564 | −7.9148 | 6.7584 | 337.8583 | 311.4491 | 3.3792 | 33.7866 | 328.42 | 17.088 | 1485 |
| 28 Dibenzo[*el* ]naphthacene | −0.8276 | −8.2948 | 7.4672 | 298.4228 | 282.1115 | 3.7336 | 29.2572 | 302.38 | 15.673 | 1062 |
| 29 Dibenzo[*de.gr* ]naphthacene | −0.8336 | −8.2774 | 7.4438 | 306.0306 | 283.3872 | 3.7219 | 29.4276 | 302.38 | 15.674 | 1110 |
| 30 Dibenzo[*g.p* ]chrysene | −0.8832 | −8.1128 | 7.2296 | 329.5913 | 310.4633 | 3.6148 | 32.2504 | 352.44 | 18.242 | 1333 |
| 31 Benzo[*c* ]picene | −0.8279 | −8.2597 | 7.4318 | 334.5104 | 311.1329 | 3.7159 | 31.7127 | 328.42 | 17.087 | 1557 |
| 32 Benzo[*ghi* ]perylene | −1.0662 | −8.0235 | 6.9573 | 273.9605 | 255.1830 | 3.4786 | 28.5940 | 276.34 | 14.259 | 815 |
| 33 Dibenzo[*b.k* ]chrysene | −1.1775 | −7.8832 | 6.7057 | 341.7884 | 312.1273 | 3.3529 | 34.0111 | 328.42 | 17.087 | 1573 |
| 34 Dibenzo[*cl* ]chrysene | −0.7827 | −8.2649 | 7.4822 | 335.5407 | 311.1978 | 3.7411 | 31.5738 | 328.42 | 17.087 | 1461 |
| 35 Benzo[*b* ]perylene | −1.1806 | −7.8666 | 6.6860 | 302.4596 | 282.8705 | 3.3430 | 31.7015 | 302.38 | 15.673 | 1088 |
| 36 Benzo[*a* ]perylene | −1.4836 | −7.5284 | 6.0448 | 302.5047 | 283.1900 | 3.0224 | 33.9068 | 302.38 | 15.673 | 1068 |
| 37 Dibenzo[*de.mn* ]naphthacene | −1.5482 | −7.4305 | 5.8823 | 306.0786 | 283.6270 | 2.9411 | 34.5268 | 302.38 | 15.673 | 1122 |
| 38 Naphtho[2.3-*g* ]chrysene | −0.9944 | −8.1177 | 7.1233 | 332.3547 | 309.9723 | 3.5617 | 32.5293 | 328.42 | 17.087 | 1405 |
| 39 Benzo[*h* ]pentaphene | −0.8089 | −8.3009 | 7.4920 | 337.8917 | 311.4461 | 3.7460 | 31.5661 | 328.42 | 17.087 | 1445 |
| 40 Benzo[*a* ]pentacene | −1.4685 | −7.5763 | 6.1078 | 345.4223 | 312.7085 | 3.0539 | 36.0786 | 328.42 | 17.087 | 1605 |
| 41 Coronene | −1.0021 | −8.1438 | 7.1417 | 289.2667 | 272.4353 | 3.5708 | 29.4287 | 300.36 | 15.413 | 1002 |
| 42 Naphtho[1.2.3.4-*def* ]chrysene | −1.0583 | −8.0233 | 6.9650 | 302.3366 | 282.8178 | 3.4825 | 30.8117 | 302.38 | 15.673 | 1082 |
| 43 Dibenzo[*def.p* ]chrysene | −1.1022 | −7.9553 | 6.8531 | 302.9476 | 282.8979 | 3.4265 | 31.1690 | 302.38 | 15.673 | 1066 |
| 44 Benzo[*rst* ]pentaphene | −1.1838 | −7.8650 | 6.6812 | 306.3398 | 283.4943 | 3.3406 | 31.7678 | 302.38 | 15.673 | 1142 |
| 45 Benzo[*g* ]chrysene | −0.7660 | −8.2705 | 7.5045 | 287.9205 | 265.8624 | 3.7522 | 27.8327 | 278.36 | 14.518 | 883 |
| 46 2.3:5.6-Dibenzophenanthrene | −0.9636 | −8.0436 | 7.0800 | 294.6767 | 266.8634 | 3.5400 | 29.1636 | 278.36 | 14.518 | 939 |
| 47 Naphtho[2.1.8-*qra* ]naphthacene | −1.3169 | −7.7299 | 6.4130 | 309.6593 | 283.9536 | 3.2065 | 32.6929 | 302.38 | 15.673 | 1166 |
| 48 Dibenz[*a.e* ]aceantrylene | −1.2806 | −8.1423 | 6.8617 | 310.6715 | 284.9750 | 3.4309 | 31.3102 | 302.38 | 15.673 | 1090 |
| 49 Acenaphthylene | −0.9359 | −8.9428 | 8.0069 | 175.2256 | 149.9876 | 4.0035 | 17.0668 | 152.20 | 7.966 | 166 |
| 50 Dibenzo[*a.k* ]fluoranthene | −1.2992 | −7.9434 | 6.6442 | 313.2985 | 285.2393 | 3.3221 | 32.0296 | 302.38 | 15.673 | 1128 |
| 51 Naphtho[2.3-*k* ]fluoranthene | −0.9127 | −7.9780 | 7.0653 | 317.5809 | 285.6889 | 3.5327 | 30.7358 | 302.38 | 15.673 | 1200 |

(*continued on next page*)

Table 2 (*continued*)

| CAS name | $E_{LUMO}$ | $E_{HOMO}$ | GAP | Area | Volume | $\eta$ | $\alpha$ | MW | $R$ | Wiener |
|---|---|---|---|---|---|---|---|---|---|---|
| 52 Fluoranthene | −0.9294 | −8.6301 | 7.7007 | 222.8304 | 195.1020 | 3.8503 | 21.5443 | 202.26 | 10.535 | 364 |
| 53 Dibenzo[*k.mno*]fluoranthene | −0.9755 | −8.4000 | 7.4245 | 280.3476 | 257.0611 | 3.7122 | 27.3474 | 276.34 | 14.259 | 848 |
| 54 1.2-Dihydroacenaphthylene | −0.2132 | −8.4944 | 8.2812 | 183.2279 | 155.6393 | 4.1406 | 16.8321 | 154.22 | 7.966 | 166 |
| 55 9H–fluorene | −0.2088 | −7.5446 | 7.3358 | 196.5286 | 166.8248 | 3.8767 | 19.1064 | 166.23 | 8.673 | 219 |
| 56 Benzo[*b*]fluorene | −0.4880 | −8.4783 | 7.9903 | 244.4588 | 212.2972 | 3.9952 | 22.1648 | 216.99 | 11.242 | 471 |
| 57 Benzo[*c*]fluorine | −0.6415 | −8.2835 | 7.6420 | 239.5300 | 211.6880 | 3.8210 | 23.0518 | 216.99 | 11.242 | 453 |
| 58 Benzo[*ghi*]fluoranthene | −0.9911 | −8.6995 | 7.7084 | 234.9511 | 211.8821 | 3.8542 | 22.8845 | 226.28 | 11.69 | 478 |
| 59 Benzo[*a*]aceanthrylene | −1.3219 | −8.0850 | 6.7631 | 266.0202 | 239.9710 | 3.3815 | 27.9729 | 252.32 | 13.104 | 666 |
| 60 Indeno[1.2.3.*cd*]pyrene | −1.2835 | −8.1363 | 6.8528 | 282.0055 | 256.9774 | 3.4264 | 29.0669 | 276.34 | 14.259 | 845 |
| 61 Indeno[1.2.3-*cd*]fluoranthene | −1.3350 | −8.5435 | 7.2085 | 286.4592 | 258.3600 | 3.6043 | 28.0860 | 313.38 | 16.121 | 871 |
| 62 Cyclopenta[*cd*]pyrene | −1.3123 | −8.2727 | 6.9604 | 234.42 | 211.87 | 3.4802 | 25.0704 | 226.28 | 11.69 | 483 |
| 63 Benzo[*j*]fluoranthene | −1.1767 | −8.3165 | 7.1398 | 265.90 | 239.95 | 3.57 | 26.7991 | 252.32 | 13.104 | 678 |
| 64 Benzo[*k*]fluoranthene | −0.9892 | −8.2994 | 7.3102 | 270.1861 | 240.39 | 3.69 | 26.0929 | 252.32 | 13.104 | 698 |
| 65 Benzo[*a*]fluorene | −0.5607 | −8.3656 | 7.8049 | 243.65 | 212.28 | 3.90 | 22.6540 | 216.28 | 11.242 | 461 |
| 66 Dibenz[*e.k*]acephenanthrylene | −1.0702 | −8.2215 | 7.1513 | 315.12 | 285.45 | 3.58 | 30.4554 | 320.38 | 14.966 | 1156 |
| 67 Benzo[*b*]fluoranthene | −0.9654 | −8.6166 | 7.6512 | 267.71 | 240.02 | 3.80 | 25.4749 | 252.32 | 13.104 | 676 |

$\eta$, Hardness; $\alpha$, polarizability; $R$, Randic connectivity index.

coefficient, $R^2 = 0.995$ and crossvalidated correlation coefficient, $Q^2 = 0.993$ for a PLS model with two latent variables (LV) describing 99.62% of total variance (Table 3). The performance of the PCR model built with the same set of descriptors was tested and one extra principal component (PC) was necessary to reach the same results obtained with the previous model. This means that all three PCs are important to build a PCR model (no data compression) and it is a consequence of the small number of variables used to model the data. A PCR model built with all the principal components (number of PCs = number of original variables) corresponds to a multiple linear regression [32]. The results obtained for these three factors in PCR model are very similar to those from PLS model: $R^2 = 0.995$ and $Q^2 = 0.994$ (Table 3).

In a previous work, Ferreira [7] obtained $R^2 = 0.999$ for a PLS model with four latent variables in a data set consisting only of 23 PAHs, all of them with just six-membered ring molecules. In the present work, a larger set of compounds including 13 molecules with five-membered ring is being

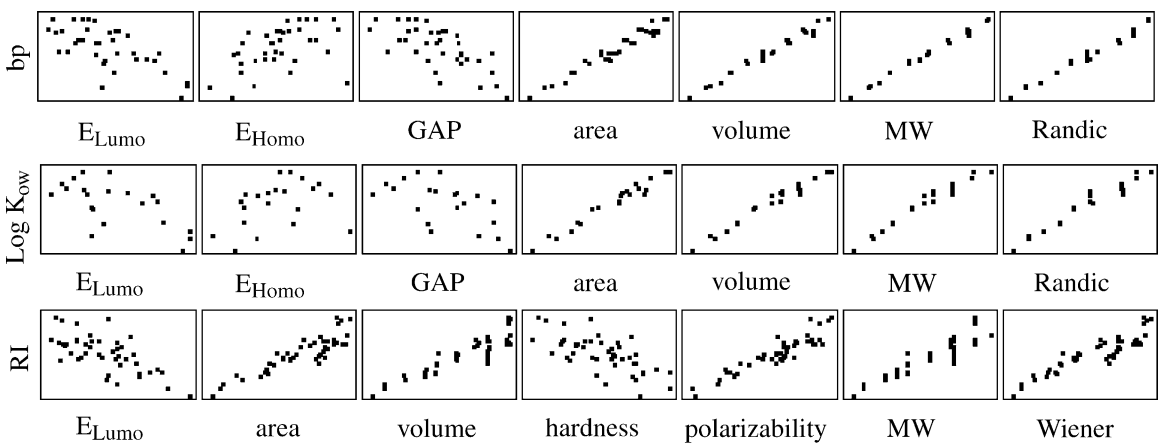


Fig. 2. Biplots showing the correlation between dependent variables, boiling point, log $K_{ow}$ and RI, and independent variables.

Table 3
QSPR results obtained by PCR and PLS for bp (°C), using the variables: volume, molecular weight (MW) and Randic connectivity index ($R$)

| | SEV | Press val | Variance (%) | $Q^2$ | $R^2$ | Variables selected | $\beta^a$ |
|---|---|---|---|---|---|---|---|
| *PCR* | | | | | | | |
| PC1 | 10.1142 | 3682.6804 | 99.5335 | 0.9894 | 0.9904 | Volume | −0.4204 |
| PC2 | 10.3821 | 3880.3789 | 99.9094 | 0.9888 | 0.9904 | MW | 1.3607 |
| **PC3** | **7.7567** | **2165.9758** | **100.0000** | **0.9938** | **0.9950** | Randic | 0.0555 |
| *PLS* | | | | | | | |
| LV1 | 10.11091 | 3680.29980 | 99.53308 | 0.9896 | 0.9904 | Volume | −0.4353 |
| **LV2** | **8.47435** | **2585.32593** | **99.62433** | **0.9942** | **0.9950** | MW | 1.3519 |
| LV3 | 7.75667 | 2165.97583 | 100.00000 | 1.0000 | 1.0000 | Randic | 0.0793 |

The models used to predicted bp are in boldcase.
[a] Regression coefficients for three principal component (PCR model) and two latent variables (PLS model).

investigated. Some of them contain $sp^3$ carbon atoms, resulting in reduced electron delocalization in the entire molecule. A larger set would be necessary to represent this class and to construct a model with a better fit, but unfortunately bp for only 14 of such structures were available in the literature. In spite of this, the results obtained were considered very satisfactory for QSPR studies. As mentioned in the methodology, the regression coefficients (Table 3) are unscaled to obtain the usual regression equations (see Appendix A for scaling details). The regression equations obtained for PCR and PLS models are described in Eqs. (7) and (8), respectively

$$\text{bp} = -22.5705 - 1.0358\text{Volume} + 2.8884\text{MW} + 2.2112R \quad (7)$$

$$\text{bp} = -21.0385 - 1.0726\text{Volume} + 2.8696\text{MW} + 3.1615R \quad (8)$$

The validation errors obtained by leave-one-out crossvalidation method for PCR and PLS models are shown in Table 4, where the first column contains the experimental values. All the bps were predicted by crossvalidation with an error lower than 4% and the predicted bps are very similar for both models. It should be emphasized that these values are obtained when the predicted compound is not included in the model. The experimental vs. predicted values using PCR and PLS models are plotted, respectively, in Fig. 3(a) and (b).

The bp of any chemical can be explained thermodynamically as the temperature at which the thermal energy of the particles is sufficient to break the cohesion forces which keep the substance in the aggregation state which characterizes the liquid phase, and allows an estimative of the atmospheric dispersion of chemicals.

The bp of PAHs can be related to the van der Waals forces for non-polar molecules, which are weaker than the dipole forces characteristic for polar molecules. The energy of these kind of intermolecular forces is closely related to structural size and molecular branching. Thus, parameters such as molecular weight, volume, surface area and Randic connectivity index are required to model bp. Although well correlated to bp, surface area (Fig. 2), was not used as a descriptor because it is highly correlated to the molecular volume, and so contains very much similar information. To illustrate the influence of the molecular size on bp, let us consider the linear polyacenes: naphthalene (bp = 218 °C; MW = 128.18), anthracene (bp = 340 °C; MW = 178.24) and naphthacene (bp = 440 °C; MW = 228.30), which differ among themselves by the number of fused rings (Fig. 1) arranged linearly. Naphthacene has two more rings than naphthalene, and its larger surface area leads to a greater number of intermolecular contacts, increasing the bp. These intermolecular van der Waals interactions are mainly of the type C–H···C–H, which are influenced by the contact area available for these interactions, expressed by the molecular weight, volume and surface area.

However, these are not the only descriptors related to the contact area. The branching effect or, in the case of PAHs, the effect of the arrangement of condensed

Table 4
Validation errors (%) for PCR (three factors) and PLS models (two factors) for boiling point (°C)

| | Experimental bp (°C) | Predicted bp | | Validation error (%) | |
|---|---|---|---|---|---|
| | | PCR model with 3 PC | PLS model with 2 LV | PCR model with 3 PC | PLS model with 2 LV |
| 1 Naphthalene | 218.00 | 225.12 | 225.08 | **− 3.27** | **− 3.25** |
| 2 Anthracene | 340.00 | 326.45 | 326.49 | **3.98** | **3.97** |
| 3 Phenanthrene | 338.00 | 327.26 | 327.34 | **3.18** | **3.15** |
| 4 Naphthacene | 440.00 | 428.48 | 426.53 | 2.62 | 3.06 |
| 5 Benz[*a* ]anthracene | 435.00 | 429.70 | 428.68 | 1.22 | 1.45 |
| 6 Chrysene | 431.00 | 430.82 | 430.51 | 0.04 | 0.11 |
| 7 Triphenylene | 429.00 | 431.79 | 431.67 | − 0.65 | − 0.62 |
| 8 Pyrene | 393.00 | 381.69 | 381.53 | 2.88 | 2.92 |
| 10 Perylene | 497.00 | 486.55 | 485.85 | 2.10 | 2.24 |
| 11 Benzo[*a* ]pyrene | 496.00 | 486.16 | 485.73 | 1.98 | 2.07 |
| 12 Benzo[*e* ]pyrene | 493.00 | 487.02 | 486.79 | 1.21 | 1.26 |
| 13 Picene | 519.00 | 537.33 | 538.86 | **− 3.53** | **− 3.83** |
| 16 Dibenz[*a.h* ]anthracene | 535.00 | 534.84 | 535.35 | 0.03 | − 0.06 |
| 17 Dibenz[*a.j* ]anthracene | 531.00 | 535.56 | 536.33 | − 0.86 | − 1.00 |
| 18 Benzo[*b* ]triphenylene | 535.00 | 535.81 | 536.05 | − 0.15 | − 0.20 |
| 23 Dibenzo[*b.def* ]chrysene | 596.00 | 588.31 | 586.75 | 1.29 | 1.55 |
| 24 Dibenzo[*def.mno* ]chrysene | 547.00 | 540.16 | 539.06 | 1.25 | 1.45 |
| 32 Benzo[*ghi* ]perylene | 542.00 | 541.16 | 540.349 | 0.15 | 0.30 |
| 41 Coronene | 590.00 | 596.80 | 595.97 | − 1.15 | − 1.01 |
| 42 Naphtho[1.2.3.4-*def* ]chrysene | 592.00 | 589.68 | 588.61 | 0.39 | 0.57 |
| 43 Dibenzo[*def.p* ]chrysene | 595.00 | 589.22 | 587.82 | 0.97 | 1.21 |
| 44 Benzo[*rst* ]pentaphene | 594.00 | 588.64 | 587.32 | 0.90 | 1.13 |
| 49 Acenaphthylene | 270.00 | 278.34 | 277.69 | − 3.09 | − 2.85 |
| 52 Fluoranthene | 383.00 | 380.79 | 380.83 | 0.58 | 0.57 |
| 54 Acenaphthene | 279.00 | 275.90 | 273.62 | 1.11 | 1.93 |
| 55 9H−fluorene | 294.00 | 302.69 | 302.76 | − 2.96 | − 2.98 |
| 56 Benzo[*b* ]fluorene | 398.00 | 406.53 | 406.95 | − 2.14 | − 2.25 |
| 57 Benzo[*c* ]fluorene | 406.00 | 406.40 | 406.76 | − 0.10 | − 0.19 |
| 58 Benzo[*ghi* ]fluoranthene | 422.00 | 437.12 | 437.01 | **− 3.58** | **− 3.56** |
| 60 Indeno[1.2.3.*cd* ]pyrene | 534.00 | 539.82 | 539.93 | − 1.09 | − 1.11 |
| 61 Indeno[1.2.3-*cd* ] fluoranthene | 531.00 | 538.37 | 538.55 | − 1.39 | − 1.42 |
| 63 Benzo[ *j* ]fluoranthene | 480.00 | 437.10 | 434.25 | − 1.52 | − 2.05 |
| 64 Benzo[*k* ]fluoranthene | 481.00 | 487.32 | 489.84 | − 1.19 | − 1.68 |
| 65 Benzo[*a* ]fluorene | 403.00 | 486.73 | 489.08 | − 1.19 | − 2.04 |
| 67 Benzo[*b* ]fluoranthene | 481.00 | 407.79 | 411.21 | − 1.28 | − 1.77 |

The validation errors were calculated from residuals (experimental−estimated value) obtained by leave-one-out crossvalidation.

rings to the bp can be understood from another four isomers with molecular weight MW = 228.30: naphthacene (bp = 440 °C), benz[*a* ]anthracene (bp = 435 °C), triphenylene (bp = 429 °C) and chrysene (bp = 431 °C), in Fig. 1. With increasing branching, the number of intermolecular contacts (C−H···C−H) tends to decrease. Therefore, the energy necessary to break these intermolecular interactions for naphthacene is greater than for triphenylene, resulting in a higher bp. The branching is, in the model, taken into account by Randic connectivity index [40,41]. The value of bp results from the influence of all the three descriptors together: molecular weight, molecular volume and Randic index.

After validation, the model constructed with the smallest validation error was used to predict the bp of the other 31 molecules. The experimental values used to build the model ranged from 218 °C (naphthalene)

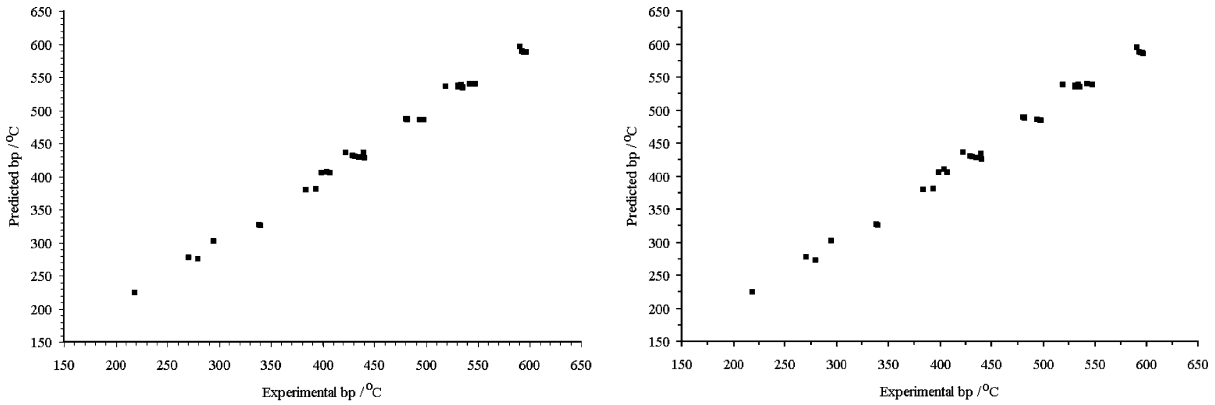


Fig. 3. Plot of experimental vs. predicted values for boiling point modeled by PCR (a) and PLS (b).

to 596 °C (dibenzo[*b.def*]chrysene). Some of the predicted values for PLS model were found beyond this range, but as tested previously, the models proposed in the present work are also robust [7]. In general, the bp predicted values obtained by both PLS and PCR models are extremely close to each other.

According to White [42] and Ferreira [7], the calculated bp for benzo[*b*]chrysene are 541 and 538 °C, respectively, which are in excellent agreement with the values predicted by the present work: 534.7 and 534.9 °C for PLS and PCR models, respectively. Their predicted bp for pentacene are 529 and 548.3 °C, respectively, while those estimated by this work are 533.3 °C (PLS) and 533.8 °C (PCR). Todeschini et al. [30] predicted a value of 297 °C for fluorene, which is very close to the values obtained in this work: 303.9 °C (PLS) and 303.3 °C (PCR).

Comparing with Ferreira's results [7], some predicted bps disagree by about 20 °C as a consequence of the extended group of molecules used in this work to construct the models, which include 13 new PAHs with five-membered rings.

### 3.2. *Octanol–water partition coefficient* ($\log K_{ow}$)

The experimental values of $\log K_{ow}$ of 21 compounds (Table 1) were used to construct the regression models. As for the bp, the same variables volume, molecular weight and Randic connectivity index [40,41] were used to build the PCR and PLS models. The regression analysis of $\log K_{ow}$ of non-substituted PAHs with these molecular descriptors resulted in a correlation coefficient $R^2 = 0.976$ and a crossvalidated correlation coefficient $Q^2 = 0.968$ with two latent variables describing 99.89% of total variance for the PLS model, and $R^2 = 0.975$, $Q^2 = 0.968$ also with two principal component describing 99.91% of total variance (Table 5) for PCR model. Ferreira [7] obtained $R^2 = 0.992$ with three latent variables for a data set consisting of PAHs with six-membered ring molecules.

The unscaled regression equation obtained for PCR model is given in Eq. (9), and the one for PLS in Eq. (10)

$$\log K_{ow} = 0.3424 + 0.0177\text{Volume} + 0.0120\text{MW} - 0.1178R \quad (9)$$

$$\log K_{ow} = 0.2231 + 0.0222\text{Volume} + 0.0084\text{MW} - 0.1230R \quad (10)$$

As shown in Table 6, except for acenaphthene, benzo[*b*]fluorene and benzo[*b*]fluoranthene both models presented validation errors lower and around 5%. In general, the PLS results are slightly better. The experimental values vs. predicted values using the PCR and PLS models are plotted, respectively, in Fig. 4(a) and (b). All the PAHs considered in this work are extremely hydrophobic, but small variations in their molecular structure can cause considerable

Table 5
QSPR results obtained by PLS and PCR for octanol–water partition coefficient (log $K_{ow}$), using the variables: volume, molecular weight (MW) and Randic connectivity index ($R$)

| | SEV (%) | Press val (%) | Variance (%) | $Q^2$ (%) | $R^2$ | Variable selected | $\beta^a$ |
|---|---|---|---|---|---|---|---|
| *PCR* | | | | | | | |
| PC1 | 0.1955 | 0.8410 | 99.5467 | 0.9630 | 0.9687 | Volume | 0.7165 |
| **PC2** | **0.1829** | **0.7355** | **99.9094** | **0.9676** | **0.9753** | MW | 0.5699 |
| PC3 | 0.1817 | 0.7262 | 100.0000 | 0.9680 | 0.9775 | Randic | −0.3012 |
| *PLS* | | | | | | | |
| LV1 | 0.19540 | 0.84003 | 99.54408 | 0.9630 | 0.9688 | Volume | 0.8983 |
| **LV2** | **0.18092** | **0.72011** | **99.88776** | **0.9683** | **0.9763** | MW | 0.4015 |
| LV3 | 0.18168 | 0.72619 | 99.99999 | 0.9680 | 0.9775 | Randic | −0.3145 |

The models used to predict log $K_{ow}$ are in boldcase.
[a] Regression coefficients for two principal component (PCR model) and two latent variables (PLS model).

changes in their affinity for organic or aqueous solvents. The values of log $K_{ow}$ are also highly related with the intermolecular forces responsible for the specific affinity of the substance with the solvent in each phase. These molecular properties are controlled by the structural size and molecular branching of the compounds.

Considering again the example illustrated previously (see structures in Fig. 1): naphthalene (log $K_{ow} = 3.37$; MW = 128.18), anthracene

Table 6
Validation errors (%) for PCR and PLS models using two factors for octanol–water partition coefficient (log $K_{ow}$)

| | Experimental log $K_{ow}$ | Predicted log $K_{ow}$ | | Validation error (%) | |
|---|---|---|---|---|---|
| | | PCR model with 2 PC | PLS model with 2 LV | PCR model with 2 PC | PLS model with 2 LV |
| 1 Naphthalene | 3.37 | 3.54 | 3.53 | **− 5.04** | −4.73 |
| 2 Anthracene | 4.54 | 4.56 | 4.56 | −0.39 | −0.43 |
| 3 Phenanthrene | 4.57 | 4.55 | 4.54 | 0.54 | 0.55 |
| 4 Naphthacene | 5.76 | 5.71 | 5.74 | 0.86 | 0.40 |
| 5 Benz[*a*]anthracene | 5.91 | 5.67 | 5.69 | 4.03 | 3.67 |
| 6 Chrysene | 5.86 | 5.67 | 5.69 | 3.19 | 2.88 |
| 7 Triphenylene | 5.49 | 5.73 | 5.74 | −4.30 | −4.54 |
| 8 Pyrene | 5.18 | 5.01 | 4.99 | 3.36 | 3.74 |
| 10 Perylene | 6.25 | 6.02 | 6.02 | 3.62 | 3.75 |
| 11 Benzo[*a*]pyrene | 6.04 | 6.06 | 6.06 | −0.36 | −0.27 |
| 16 Dibenz[*a.h*]anthracene | 6.75 | 6.75 | 6.78 | −0.01 | −0.51 |
| 32 Benzo[*ghi*]perylene | 6.5 | 6.53 | 6.50 | −0.43 | −0.06 |
| 41 Coronene | 6.75 | 7.05 | 7.04 | −4.50 | −4.34 |
| 49 Acenaphthylene | 4 | 3.99 | 3.97 | 0.15 | 0.69 |
| 52 Fluoranthene | 5.22 | 5.03 | 5.02 | 3.59 | 3.78 |
| 54 Acenaphthene | 3.92 | 4.17 | 4.17 | **− 6.38** | **− 6.31** |
| 55 9H–fluorene | 4.18 | 4.36 | 4.36 | −4.36 | −4.42 |
| 56 Benzo[*b*]fluorene | 5.75 | 5.38 | 5.39 | **6.46** | **6.31** |
| 64 Benzo[*k*]fluoranthene | 6 | 6.09 | 6.10 | −1.56 | −1.59 |
| 65 Benzo[*a*]fluorene | 5.4 | 5.37 | 5.38 | 0.56 | 0.31 |
| 67 Benzo[*b*]fluoranthene | 5.8 | 6.12 | 6.11 | **− 5.47** | **− 5.43** |

The validation errors were calculated from residuals (experimental–estimated value) obtained by leave-one-out crossvalidation.

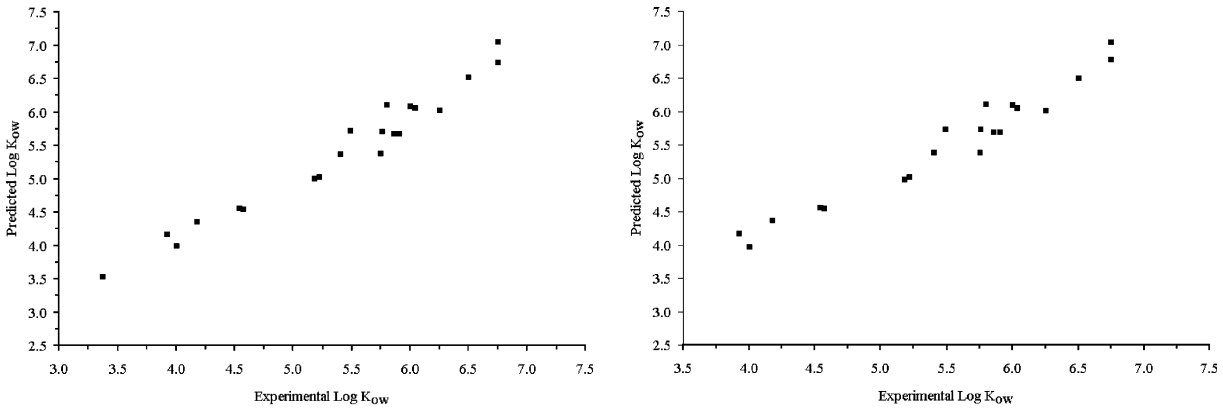


Fig. 4. Plot of experimental vs. predicted values for log $K_{ow}$ modeled by PCR (a) and PLS (b).

($\log K_{ow} = 4.54$; MW = 178.24) and naphthacene ($\log K_{ow} = 5.76$; MW = 228.30), the increase in hydrophobicity (described by $\log K_{ow}$ values) as a function of the number of rings and molecular size is clearly seen.

The importance of the arrangement of condensed rings can be discussed considering the four isomer structures (MW = 228.30): naphthacene ($\log K_{ow} = 5.76$), benz[*a*]anthracene ($\log K_{ow} = 5.91$), triphenylene ($\log K_{ow} = 5.49$) and chrysene ($\log K_{ow} = 5.86$). The spatial distribution of rings in these molecules results in small branching, which is described in the constructed models by molecular volume. Benzo[*b*]fluorene is one of the few compounds with five-membered ring (see structures in Fig. 1) and lower electron delocalization. All these five-membered ring molecules exhibit a planar structure, having just the hydrogen atoms bonded to the $sp^3$ carbon atom, out of the molecular plane. For the bp, all of them (acenaphthene, fluorene, benzo[*b*]fluorene, benzo[*c*]fluorene and benzo[*a*]-fluorene) were included to construct the model, resulting in a more representative calibration set for these PAHs. For $K_{ow}$, there is no experimental data for benzo[*c*]fluorene, resulting in larger validation errors for acenaphthene, and particularly for benzo[*b*]fluorene, because this class has only a few representative members when constructing the models and two of them are isomers.

Benzo[*b*]fluoranthene, another structure with a high validation error, belongs to a small group of PAHs with six- and five-membered rings, in which there are bonds that are always single (see the structures in Fig. 1). To construct the bp model, six of these structures were used and for the log $K_{ow}$ model just three. This could be the source of the increased validation error for benzo[*b*]fluoranthene, although the error could be originated from uncertainties in experimental determinations or other unknown factors. It should be emphasized that this discussion is based on validation errors, originated from prediction when the compound is not in the model, and so, different from the predicted value using Eqs. (10) and (11), where these errors are smaller.

After validation, the two-factor models were used to predict the log $K_{ow}$ of the other 46 molecules. Table 1 shows the experimental and predicted log $K_{ow}$ values. The experimental values used to build the model were in the range of 3.37 (naphthalene) to 6.75 (dibenz[*a.h*]anthracene and coronene), and some of the predicted values were found beyond this range but very close to the extreme values.

Todeschini et al. [30] and Ferreira [7] predicted the log $K_{ow}$ for dibenz[*a,j*]anthracene as 6.62 and 6.87, respectively, which are in excellent agreement with the predicted values obtained in the present work: 6.67 for PLS and 6.75 for PCR models. Todeschini et al. [30] calculated 6.07 to the log $K_{ow}$ value for benzo[*ghi*]fluoranthene, which are in good agreement with values obtained in this work: 5.49 (by PLS) and 5.51 (by PCR).

### 3.3. Retention time index for reversed-phase LC analysis

The regression models for RI were constructed using experimental data from a group of 44 molecules of non-substituted PAHs (Table 1). The molecular descriptors $E_{LUMO}$, hardness ($\eta$), polarizability ($\alpha$), molecular weight (MW) and Wiener index (Wiener) [5] listed in Table 2 were selected to build the model.

The regression analysis of RI with these molecular descriptors resulted in a correlation coefficient $R^2 = 0.898$ and a crossvalidated correlation coefficient $Q^2 = 0.882$ for the PLS model with two latent variables describing 97.53% of total variance. The PCR model using also two principal components resulted in $R^2 = 0.898$ and $Q^2 = 0.882$, for 97.63% of total variance explained (Table 7). Ferreira [7] obtained $R^2 = 0.9702$ for a model with three latent variables in a data set consisting only of 33 PAHs, all of them with six-membered ring. The unscaled regression equations obtained for PCR and PLS models, using the regression coefficients in Table 7, are described in Eqs. (11) and (12), respectively

$$\mathrm{RI} = 2.3164 - 0.3308E_{\mathrm{LUMO}} - 0.3588\eta + 0.0470\alpha + 0.0049\mathrm{MW} + 0.0007\mathrm{Wiener} \quad (11)$$

$$\mathrm{RI} = 2.3561 - 0.3289E_{\mathrm{LUMO}} - 0.3627\eta + 0.0465\alpha + 0.0048\mathrm{MW} + 0.0007\mathrm{Wiener} \quad (12)$$

Although the plots of experimental vs. predicted values (Fig. 5(a) and (b)) show a larger spread than expected, the validation errors are very satisfactory for QSPR studies, with residuals smaller than 10% (Table 8), where only benzo[*b*]fluorene gave a validation error above this limit (10.44%).

Again, the validation errors obtained by the PLS model were smaller than those obtained by the PCR model. The experimental values of RI used to construct the models are between two (naphthalene) and six (dibenz[*a.h*]anthracene and coronene).

RI is a parameter used in chromatographic analysis to separate hydrocarbon mixtures, and provide the elution order of each compound [15]. This order is a result of complex factors, and involves interactions between the mobile and stationary phase. These interactions are very much related to the molecular structure of both phases, and the variations in RI result from the structural variations of the compounds.

For instance, larger compounds tend to have higher retention times, as they possess larger surface areas that promote the formation of stronger interactions with the stationary phase in comparison to smaller compounds. These interactions hinder the diffusion of the solute through the stationary phase, increasing the retention time of the compound. Compare the RI of anthracene (RI = 3.2) and naphthacene, RI = 4.51).

The strength of these intermolecular interactions also depends on factors such as the number of rings

Table 7
QSPR results obtained by PLS and PCR for retention time index (RI), using the variables: LUMO's Energy ($E_{LUMO}$), hardness ($\eta$), polarizability ($\alpha$), molecular weight (MW) and Wiener index

| | SEV (%) | Press val (%) | Variance (%) | $Q^2$ (%) | $R^2$ (%) | variables selected | $\beta^a$ |
|---|---|---|---|---|---|---|---|
| *PCR* | | | | | | | |
| PC1 | 0.2529 | 2.8142 | 81.9049 | 0.8729 | 0.8861 | $E_{LUMO}$ | −0.1329 |
| *PC2* | **0.2435** | **2.6080** | **97.6292** | **0.8822** | **0.8980** | $\eta$ | −0.1232 |
| PC3 | 0.2494 | 2.7358 | 99.6965 | 0.8764 | 0.8981 | $\alpha$ | 0.2416 |
| PC4 | 0.2525 | 2.8046 | 99.9492 | 0.8733 | 0.9009 | MW | 0.2644 |
| | | | | | | Wiener | 0.2654 |
| *PLS* | | | | | | | |
| LV1 | 0.2493 | 2.7343 | 81.7339 | 0.8765 | 0.8903 | $E_{LUMO}$ | −0.1321 |
| *LV2* | **0.2442** | **2.6228** | **97.5313** | **0.8815** | **0.8981** | $\eta$ | −0.1246 |
| LV3 | 0.2624 | 3.0300 | 97.9662 | 0.8632 | 0.9004 | $\alpha$ | 0.2387 |
| LV4 | 0.2537 | 2.8315 | 99.9243 | 0.8721 | 0.9011 | MW | 0.2595 |
| | | | | | | Wiener | 0.2729 |

The models used to predict the retention index are in boldcase.
[a] Regression coefficients for two principal component (PCR model) and two latent variables (PLS model).

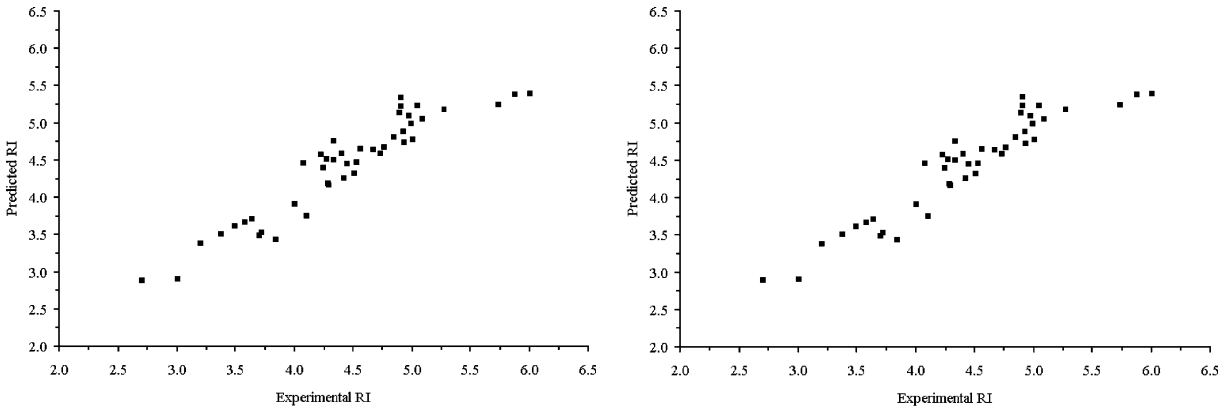


Fig. 5. Plot of experimental vs. predicted values for retention time index (RI) modeled by PCR (a) and PLS (b).

and the branching. For example, in a set of isomers with the same molecular weight (MW = 228.3), the larger the volume and surface area, the greater the RI. Compare naphtacene (RI = 4.51), benz[*a*]anthracene (RI = 4.00), chrysene (RI = 4.10) triphenylene (RI = 3.7) and benzo[*c*]phenanthrene (3.64). The RI decreases for more compact molecules. But there is another parameter to be considered: the planarity. Benzo[*c*]phenanthrene is the only compound among these four which shows a non-planar and helicoidal structure (Fig. 6) and this peculiarity influences their elution time. Another example on the effect of this non-planarity which occurs due to the steric repulsion between the hydrogen atoms from proximal rings in 3D space, and influences the elution time, is given by dibenz[*a.j*]anthracene (RI = 4.56), benzo[*c*]chrysene (RI = 4.45) and dibenzo[*g.c*]phenanthrene (RI = 4.07) [15]. This information about compactness, ring distribution, number of rings and planarity is provided in Eqs. (11) and (12) by the topological descriptor named Wiener index [5,25]. This topological index was proposed by Wiener in 1947 and describes the molecular ramifications through a distance matrix. The matrix elements are the distances between pairs of atoms (given by the smallest number of chemical bonds between them) in the molecule. The Wiener index is inversely proportional to the compactness of the molecule [5,25]. Besides the molecular weight and the Wiener index, three electronic variables contributed to the construction of the regression model: the energy of the frontier orbital LUMO, hardness and polarizability.

Hardness is defined by the gap energy between the frontier orbitals, described in Eq. (13) where $\eta$ is the hardness, and $E_{\mathrm{LUMO}}$ and $E_{\mathrm{HOMO}}$ are, respectively, the LUMOs and HOMOs energies

$$\eta = \tfrac{1}{2}(E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}) \tag{13}$$

The absolute hardness is a measure of energy stabilization of a system. The greater the hardness, the greater the stability of a chemical. Polarizability ($\alpha$), indicates the ease with which species can be deformed by an electric field. One of the most important polarizing influences on a chemical species is that arising from the presence of another species nearby. The changes in the wavefunction and energies are responsible for the adhesion of both species, resulting in intermolecular interactions. These intermolecular forces are directly related to the RI values.

In spite of the correlation among these variables, all of them give some complementary information, essential for the good fitting of the model. Some models were constructed excluding each electronic variable, one at a time, however, the obtained results were inferior to those obtained when the three of them were included simultaneously.

After validation, the model constructed was used to predict the RI of 23 other PAHs. Just one predicted value was found outside the range of the modeling set, but very close to the extreme (Table 1) and the values

Table 8
Validation errors (%) for PCR and PLS models using two factors for retention time index (RI)

| | Experimental RI | Predicted RI | | Validation Error (%) | |
|---|---|---|---|---|---|
| | | PCR model with 2 PC | PLS model with 2 LV | PCR model with 2 PC | PLS model with 2 LV |
| 2 Anthracene | 3.2 | 3.38 | 3.39 | −5.75 | −5.79 |
| 3 Phenanthrene | 3 | 2.91 | 2.91 | 3.08 | 3.04 |
| 4 Naphthacene | 4.51 | 4.33 | 4.33 | 4.00 | 4.03 |
| 5 Benz[*a* ]anthracene | 4 | 3.91 | 3.91 | 2.18 | 2.19 |
| 6 Chrysene | 4.1 | 3.75 | 3.75 | 8.43 | 8.43 |
| 7 Triphenylene | 3.7 | 3.49 | 3.49 | 5.55 | 5.58 |
| 8 Pyrene | 3.58 | 3.67 | 3.67 | −2.64 | −2.62 |
| 9 Benzo[*c* ]phenanthrene | 3.64 | 3.72 | 3.72 | −2.21 | −2.20 |
| 10 Perylene | 4.33 | 4.51 | 4.51 | −4.20 | −4.17 |
| 11 Benzo[*a* ]pyrene | 4.53 | 4.47 | 4.47 | 1.26 | 1.28 |
| 12 Benzo[*e* ]pyrene | 4.28 | 4.19 | 4.19 | 2.14 | 2.17 |
| 14 Pentaphene | 4.67 | 4.65 | 4.65 | 0.49 | 0.46 |
| 15 Benzo[*b* ]chrysene | 5 | 4.78 | 4.78 | 4.34 | 4.33 |
| 16 Dibenz[*a.h* ]anthracene | 4.73 | 4.59 | 4.59 | 2.91 | 2.88 |
| 17 Dibenz[*a.j* ]anthracene | 4.56 | 4.66 | 4.66 | −2.16 | −2.18 |
| 18 Benzo[*b* ]triphenylene | 4.4 | 4.59 | 4.59 | −4.31 | −4.31 |
| 19 Benzo[*c* ]chrysene | 4.45 | 4.46 | 4.46 | −0.20 | −0.22 |
| 21 Dibenzo[*c.g* ]phenanthrene | 4.07 | 4.47 | 4.47 | −9.79 | −9.80 |
| 22 Benzo[*a* ]naphthacene | 4.99 | 5.00 | 5.00 | −0.17 | −0.19 |
| 23 Dibenzo[*b.def* ]chrysene | 6 | 5.40 | 5.40 | 9.94 | 9.96 |
| 24 Dibenzo[*def.mno* ]chrysene | 5.08 | 5.06 | 5.06 | 0.46 | 0.49 |
| 29 Dibenzo[*de.gr* ]naphthacene | 4.92 | 4.89 | 4.89 | 0.68 | 0.66 |
| 32 Benzo[*ghi* ]perylene | 4.76 | 4.68 | 4.68 | 1.65 | 1.68 |
| 35 Benzo[*b* ]perylene | 5.04 | 5.24 | 5.24 | −4.03 | −4.05 |
| 42 Naphtho[1.2.3.4-*def* ]chrysene | 4.97 | 5.10 | 5.10 | −2.67 | −2.68 |
| 43 Dibenzo[*def.p* ]chrysene | 4.89 | 5.15 | 5.15 | −5.30 | −5.30 |
| 44 Benzo[*rst* ]pentaphene | 5.73 | 5.25 | 5.25 | 8.45 | 8.45 |
| 45 Benzo[*g* ]chrysene | 4.27 | 4.52 | 4.52 | −5.88 | −5.88 |
| 46 2.3:5.6-Dibenzophenanthrene | 4.33 | 4.76 | 4.76 | −10.00 | −10.03 |
| 47 Naphtho[2.1.8-*qra* ]naphthacene | 5.87 | 5.39 | 5.39 | 8.20 | 8.21 |
| 48 Dibenz[*a.e* ]aceantrylene | 4.9 | 5.23 | 5.24 | −6.83 | −6.85 |
| 50 Dibenzo[*a.k* ]fluoranthene | 4.9 | 5.35 | 5.35 | −9.20 | −9.22 |
| 52 Fluoranthene | 3.37 | 3.51 | 3.51 | −4.25 | −4.30 |
| 55 9H−fluorene | 2.7 | 2.89 | 2.90 | −7.00 | −7.41 |
| 56 Benzo[*b* ]fluorene | 3.84 | 3.44 | 3.44 | **10.44** | **10.44** |
| 57 Benzo[*c* ]fluorene | 3.49 | 3.62 | 3.62 | −3.81 | −3.80 |
| 59 Benzo[*a* ]fluoranthene | 4.22 | 4.59 | 4.59 | −8.66 | −8.66 |
| 60 Indeno[1.2.3.*cd* ]pyrene | 4.84 | 4.82 | 4.82 | 0.49 | 0.52 |
| 61 Indeno[1.2.3-*cd* ] fluoranthene | 4.93 | 4.74 | 4.73 | 3.90 | 4.06 |
| 63 Benzo[ *j* ]fluoranthene | 4.24 | 4.40 | 4.40 | −3.76 | −3.76 |
| 64 Benzo[*k* ]fluoranthene | 4.42 | 4.26 | 4.26 | 3.53 | 3.56 |
| 65 Benzo[*a* ]fluorene | 3.72 | 3.54 | 3.54 | 4.93 | 4.94 |
| 66 Dibenz[*e.k* ]acephenanthrylene | 5.27 | 5.18 | 5.18 | 1.67 | 1.67 |
| 67 Benzo[*b* ]fluoranthene | 4.29 | 4.17 | 4.17 | 2.73 | 2.80 |

The validation errors were calculated from residuals (experimental−estimated value) obtained by leave-one-out crossvalidation.

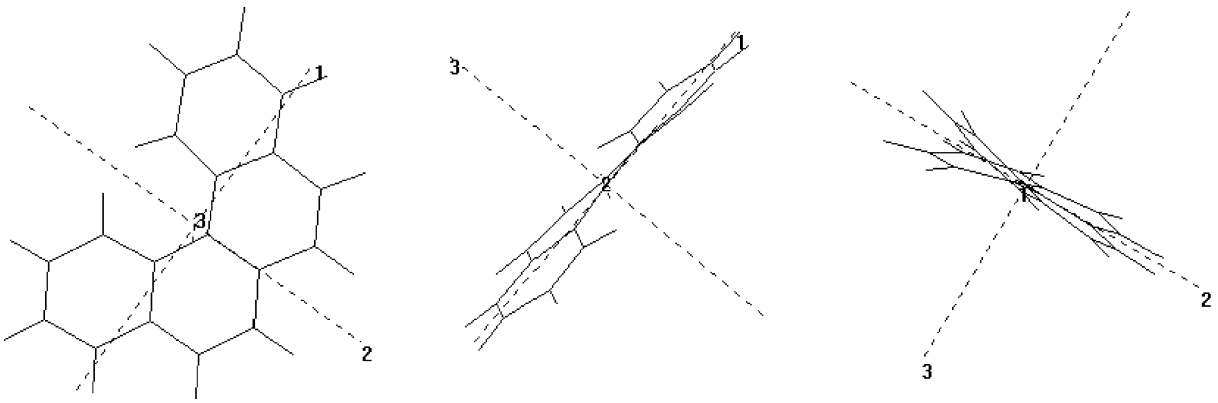


Fig. 6. Spacial rotation of benzo[*c*]phenanthrene.

obtained by PCR and PLS are almost identical. Ferreira [7] predicted RI for pentacene, dibenzo[*a,j*]-naphthacene, dibenzo[*a,l*]naphthacene and coronene as 5.72, 6.30, 6.27 and 5.21, respectively, which are in excellent agreement with the predicted values of 5.40, 5.77, 5.75 and 4.91 obtained in the present work using PLS model. The values obtained for these compounds by PCR model are, respectively, 5.40, 5.76, 5.75 and 4.92.

## 4. Conclusions

QSPR studies are an important tool for research and knowledge of chemical compounds and it has been frequently used in medicinal chemistry and molecular design to investigate new drugs. It is especially useful when the experimental determination of properties is very complex, the handling of materials may involve some risk, or determinations may not be easy in cases where compounds can quickly degrade. In general, the experimental determinations are very expensive and the QSPR studies allow a reduction of this cost. Experimental difficulties in the determination of the bp and $\log K_{ow}$ of PAHs arise from the fact that besides their high toxicity, they are highly hydrophobic.

From a set of 168 molecular descriptors, it was possible to select a few of them to construct regression models with good predictability for three physicochemical properties: bp, $\log K_{ow}$ and RI. The PLS method has been shown to be a useful and powerful tool to construct these QSPR models and allow the prediction of the properties not available yet in the literature for five- and six-membered rings PAHs. PCR, another multivariate repression method not so popular in QSPR studies was also tested for comparison. It could be observed that the statistical parameters obtained for both models PCR/PLS are very close to each other. The same was found for the predicted values of bp, $K_{ow}$ and RI.

## Acknowledgements

This work was supported by agencies FAPESP, CAPES and FAEP. We thank Dr Lucicleide R. Cirino for assistance with the theoretical calculations and valuable suggestions, and Dr Stephen Gurden for the text revision. Also the National Center for High Performance Computing in São Paulo (CENAPAD), for computational support.

## Appendix A. Calculations of the unscaled regression vector (from autoscaled data) using MATLAB code

For each calibration data set:

(1) denote the data matrix by **X**;
(2) denote the dependent variable vector by **y**;
(3) denote the autoscaled regression vector (from the

model) by **beta**;
(4) calculate the mean of **y** by $ymean = mean(y)$;
(5) calculate the mean of **X** by $xmean = mean(X)$;
(6) calculate the standard deviation of $y$ by $ystd = std(y)$;
(7) calculate the standard deviation of $X$ by $xstd = std(X)$;
(8) find the unscaled regression vector calculating the indices $\beta_0$ and $\beta_1$, where $N$ is the number of independent variables used in the model:

$$\beta_0 = (ymean - ((xmean. * (ones(1, N) * ystd./xstd)) * beta))$$

$$\beta_1 = (diag(ones(1, N) * ystd./xstd)) * beta$$

(9) the regression equation is: $y = \beta_0 + (\beta_1 * X)$;
(10) check the validation of the equation calculating the calibration errors (use the calibration data set)

$$y_{pred} = \beta_0 + (\beta_1 * X)$$

$$Errors = y - y_{pred}$$

For the prediction data set

(1) denote the data matrix by **X**;
(2) denote the dependent variable vector by **y**;
(3) apply this new equation to find the predicted values of **y** for the prediction data set.

## References


[1] D. Mackay, W.Y. Shiu, K.C. Ma, Illustrated Handbook of Physical–Chemical Properties and Environmental Fate for Organic Chemicals, vol. 3, Lewis, London, 1998.
[2] H. Kubinyi, QSAR: Hansch Analysis and Related Approaches, VCH, Weinheim, 1993.
[3] M. Karelson, A. Perkson, Comput. Chem. 23 (1999) 49.
[4] S. Arupjyoti, S. Iragavarapu, Comput. Chem. 22 (1998) 515.
[5] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17.
[6] H.B. Krop, J.M. van Velzen, J.J.R. Parsons, H.A.J. Goves, Chemosphere 34 (1997) 107.
[7] M.M.C. Ferreira, Chemosphere 44 (2001) 125.
[8] L. Cirino, M.M.C. Ferreira, Quím. Nova 26 (2003) 312.
[9] L.P. Burkhard, A.W. Andren, D.E. Armstrong, Environ. Sci. Technol. 19 (1985) 500.
[10] E.U. Ramos, W.H.J. Vaes, H.J.M. Verhaar, J.L. Hermes, J. Chem. Inf. Comput. Sci. 38 (1998) 845.
[11] Y.H. Zhao, G.D. Ji, M.T.D. Cronin, J.C. Dearden, Sci. Total Environ. 216 (1998) 205.
[12] H.J.M. Verhaar, E.U. Ramos, J.L. Hermens, J. Chemom. 10 (1996) 149.
[13] M.T.D. Cronin, T.W. Schultz, Sci. Total Environ. (1997) 75.
[14] A.D.P. Netto, J.C. Moreira, A.E.X.O. Dias, G. Arbilla, L.F.V. Ferreira, A.S. Oliveira, J. Barek, Quím. Nova 23 (2000) 765.
[15] L.C. Sander, S.A. Wise, Adv. Chromatogr. 25 (1986) 139.
[16] A.F. Lehner, J. Horn, J.W. Flesher, J. Mol. Struct. 366 (1996) 203.
[17] S.G. Huling, J.W. Weaver, Dense nonaqueous phase liquids. Ground water issue. EPA/540/4-91-002, US EPA. R.S. Kerr Environmental Research Laboratory, Ada, OK, p. 21.
[18] C.J. Newell, S.D. Acree, R.R. Ross, S.G. Huling, Light nonaqueous phase liquids. Ground water issue. EPA/540/S-95/500, US EPA. R.S. Kerr Environmental Research Laboratory, Ada, OK, p. 28.
[19] B. Elvers, S. Harokins Ullman's Encyclopedia of Industrial Chemistry, vol. B7, fifth ed., VCH, New York, 1996.
[20] W. Karcher, R.J. Fordham, J.J. Dubois, P.G.J.M. Glaude, J.A.M. Lighart, Spectral Atlas of Polycyclic Aromatic Compounds, vol. 2, Kluwer Academic Publishers, Dordrecht, 1988.
[21] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[22] PC SPARTAN® 1.0.5., Wavefunction, Inc., Irvine, CA, 2001.
[23] M. Huang, D. Doerge, N. Bodor, E. Pop, M.E. Brewster, Int. J. Quantum Chem.: Quantum Biol. Symp. 22 (1995) 171.
[24] N.S. Bodor, M. Huang, Int. J. Quantum Chem.: Quantum Biol. Symp. 61 (1997) 127.
[25] H. Van der Waterbeemd, Chemometric Methods in Molecular Design, vol. 2, VCH, Weinheim, 1995.
[26] P. Labute, J. Mol. Graphics Modelling 18 (2000) 464.
[27] W.J. Hehre, L.D. Burke, A.J. Shusterman, W.J. Pietro, Experiments in Computational Organic Chemistry, Wavefunction, Irvine, CA, 1993.
[28] F. Jensen, Introduction to Computational Chemistry, Wiley, New York, 1999.
[29] R. Todeschini, M. Lasagni, E. Marengo, J. Chemom. 8 (1994) 263.
[30] R. Todeschini, P. Gramatica, R. Provenzani, E. Marengo, Chemom. Intell. Lab. Syst. 27 (1995) 221.
[31] R. Todeschini, P. Gramatica, Quant. Struct.-Act. Relat. 16 (1997) 113.
[32] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1.
[33] P. Geladi, J. Chemom. 2 (1988) 231.
[34] M.M.C. Ferreira, J. Braz. Chem. Soc. 13 (2002) 742.
[35] R. Custódio, J.C. Andrade, F. Augusto, Quím. Nova 20 (1997) 219.
[36] M. Pimentel, B.B. Neto, Quím. Nova 19 (1996) 268.
[37] Pirouette, Multivariate Data analysis for IBM PC Systems, 3, 01, Infometrix, Inc., Woodinville, WA, 2001.
[38] PLS_Toolbox 2.0. Eigenvector Research: Manson, WA.
[39] MATLAB®. for Windows, version 5.0, MathWorks, Inc., 1995.
[40] M. Randic, Chem. Phys. Lett. 211 (1993) 478.
[41] M. Randic, J. Am. Chem. Soc. 97 (1975) 6609.
[42] C.M. White, J. Chem. Engng Data 31 (1986) 198.