Journal of Chromatography A, 1233 (2012) 116–125

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Investigation of retention behavior of polychlorinated biphenyl congeners on 18 different HRGC columns using molecular surface average local ionization energy descriptors

Raouf Ghavami*, Bakhtyar Sepehri

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

ARTICLE INFO

Article history:
Received 7 December 2011
Received in revised form 16 January 2012
Accepted 17 January 2012
Available online 11 February 2012

Keywords:
Relative retention time
Polychlorinated biphenyl
Molecular surface electrostatic potentials
GIPF approach
Molecular surface average local ionization energy

ABSTRACT

In this paper, based on the general interaction properties function (GIPF) family descriptors computed at the B3LYP/6-31G* level in Gaussian98 software, a significant quantitative structure–retention relationship (QSRR) models for the high resolution gas chromatographic relative retention time (HRGC-RRT) of all PCB congeners on 18 different HRGC capillary columns were constructed by using multiple linear regression (MLR) analysis, following the guidelines for development and validation of QSRR models. By means of the elimination selection stepwise regression algorithms, the molecular surface average local ionization energy was selected as one-parameter univariate linear regression to develop a QSRR model for prediction of GC-RRT of PCBs on each stationary phase. The accuracy of all developed models was confirmed using different types of internal and external procedures. A successful interpretation of the complex relationship between HRGC-RRTs of PCBs and the chemical structures was achieved by QSRR.

1. Introduction

Polychlorinated biphenyls (PCBs) are a class of biphenyl compounds with one to ten hydrogen atoms replaced by chlorine. At room temperature, they range in physical state from light- to dark-yellow oily liquids to white crystalline solids and hard non-crystalline resins [1,2]. PCBs are produced commercially, depending upon the number of chlorines and their location on the biphenyl rings, resulting in 209 possible PCB congeners [3]. The name of a congener specifies the total number of chlorine substitutions and the position of chlorine. Fig. 1 shows the conventional nomenclature, with the 1 and 6 positions closest to the biphenyl bond described as *ortho* (carbon atoms 2, 2′, 6, or 6′), those opposite called *para* (carbon atoms 4 or 4′), and the remainder called *meta* (carbon atoms 3, 3′, 5, or 5′). The number and position of the chlorine atoms determine both the physical and the biological properties of each PCBs [4]. PCBs with fewer chlorine atoms tend to be more soluble in water, more volatile, and more easily metabolized. Larger numbers of chlorine atoms are associated with increased resistance to biodegradation, which can increase bioaccumulation in the environment. PCBs are practically insoluble in water, but soluble in organic solvents and fats. They are very stable and persistent in the environment [5]. PCBs were used widely in electrical equipment like capacitors and transformers. They also were used in hydraulic fluids, heat transfer fluids, lubricants, and plasticizers. PCBs are persistent environmental pollutants that elicit a number of adverse health effects including teratogenesis, neurotoxicity, immunotoxicity, reproductive toxicity, endocrine disruption and carcinogenesis [6]. The major emission of PCBs in the environment has been related to the use of PCB mixtures in electrical equipment, but coal combustion, steel melting and waste incineration are also among the possible sources. Although PCBs were banned between the 1970s and 1980s in most industrialized countries because of their alleged carcinogenicity, they are still present in the environment. Depending on the structural characteristics, such as the specific pattern of chlorine substitution in *para* and *meta* or *ortho* positions, PCBs are classified as non-*ortho*-substituted dioxin-like (DL) or *ortho*-substituted non-dioxin-like (NDL) congeners. DL-PCBs have high affinity for the aryl hydrocarbon receptor (AhR), a ligand-activated transcription factor that controls the expression of cytochrome P450 1A (CYP1A) genes and are regarded to be highly toxic, similar to those of the toxic effects of TSDD (2,3,7,8-tetrachlorodibenzo-*p*-dioxin). NDL-PCB congeners are generally considered less toxic, but the nervous system has appeared to be one of their most sensitive targets [4,7–9]. Association between elevated exposure to PCB mixtures and alterations in liver enzymes, hepatomegaly, and dermatological effects such as rashes and acne has been reported [10].

* Corresponding author. Tel.: +98 871 6624133; fax: +98 871 6624133.
E-mail addresses: r.ghavami@uok.ac.ir, rghavami2000@yahoo.com (R. Ghavami).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2012.01.047

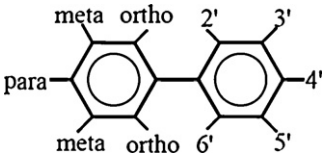

Fig. 1. Chemical structure of the PCB molecule. The 10 positions are numbered 2–6 on one ring and 2′–6′ on the second ring. Chlorines can be substituted for hydrogen at 10 possible sites on the biphenyl rings.

Due to PCBs' complex composition, many researchers have also placed emphasis on the identification of the individual PCB congeners [11,12]. Presently, all 209 PCB congeners have been commercially synthesized and are available for use as standards, and because of advances in high resolution gas chromatography (HRGC), it is possible to determine most of the individual PCB congeners in the environmental samples. However, separation and characterization of all 209 PCB congeners is still an extremely difficult (if not impossible) task that attracts on-going research focus in this field [13]. One of the most successful approaches to the prediction of physicochemical and biological properties of organic molecules, starting only from molecular structure information, is quantitative structure–property/activity relationships modeling (QSPRs/QSARs) [14]. QSPRs/QSARs are mathematical models that attempt to correlate the molecular structure of compounds and their biological, chemical, and physical properties. Among the most extensively studied properties are the chromatographic ones. It is considered that the same basic intermolecular interactions determine the behavior of chemical compounds in both biological and chromatographic environments [15]. Retention in chromatography is the result of a competitive distribution process of the solute between mobile and stationary phases, in which, the partitioning of the solute between these phases is largely determined by the molecular structure [16]. Predicting chromatographic behavior from molecular structure of solutes resulted in the quantitative structure–retention relationships (QSRRs) methodology. QSRRs are statistically derived relationships between the chromatographic parameters determined for a representative series of analytes in given separation systems and the molecular descriptors accounting for the structural differences among the investigated analytes [17]. Such relationships may provide insight into the molecular mechanism of separation in a given chromatographic system, generate knowledge about the various interactions taking place between the solute and the stationary phase, evaluate physicochemical properties of analytes and identify the most informative structural descriptors. Due to the need to control the PCBs level in the environment, the analytical methods for their analysis are currently based on their separation by GC [18] using on capillary columns with different polarities [19,20] and specific detectors such as the flame ionization detector (FID) [21], the electron-capture detector (ECD) [22,23] and mass spectrometry (MS) [24,25]. The HRGC-RRT on a capillary column with detection by ECD is a unique characteristic of the PCBs and can be used for the identification purpose. Despite the broad range of GC stationary phases available, none can separate all PCBs from each other. Various techniques have been used based on the combination of commercially available GC columns to improve the separation efficiency of PCBs [26–29]. At present, a database of RRTs and coelutions for all 209 congeners on 20 different stationary phases with MS or ECD detection has been reported [12,30]. In the past, several attempts have been made to build QSRR models on the prediction of RRT for PCBs on different stationary phases [16,31–36]. Hasan and Jurs [32] used the five-variable regression equation for prediction of GC-RRT of 209 PCBs with $R^2 = 0.997$ and standard deviation of 0.017. Liu et al. [33] used the five-variable regression equation with $R^2 = 0.9928$ and the root mean square errors of 0.0152 based on molecular electronegativity distance vector (MEDV) descriptors to correlate with the GC-RRTs of 209 PCB congeners on the SE-54 stationary phase. Ren et al. [35] using CODESSA software package and principal component analyzed (PCA) presented a QSRR study for the GC–GC–TOFMS (time-flight mass spectrometry) chromatographic relative retention time of 209 PCB congeners. PCA was used to recognize groups of samples with similar behavior and assist the separation of the data into training and test sets. Jäntschi et al. [36] reported the use of a molecular descriptors family (MDF) in QSRR modeling to predict the chromatographic relative retention times of 209 PCBs on a capillary column of SE-54.

In two previous researches, we create QSRR models on 18 stationary phases using Dragon descriptors and number of chlorine atoms in different positions on two phenyl rings as descriptors [37,38]. These QSRR models can detect the best column for separation of PCBs but recognition of separation mechanism is not easy because interpretation of these descriptors is problem. The object of the present investigation, was to find QSRR models, for try to solve this problem and with the best predictive and explanatory performance, for HRGC-RRTs of 209 PCBs congener's values on 18 stationary phases by using molecular descriptors derived from molecular surface electrostatic potentials and average local ionization energy. Retention is a phenomenon primarily dependent on the interactions between the solute and the stationary-phase molecules such as induction force, dispersion force, and hydrogen bond that related to the topological structures, geometric and electronic environments of the solute [39]. It has long been recognized that non-covalent interactions are predominantly electrostatic in nature. Politzer et al. have shown that a variety of condensed phase macroscopic properties that depend on non-covalent interactions can be expressed analytically in terms of statistically defined quantities that characterize molecular surface electrostatic potentials and average local ionization energy and named their models, general interaction properties function (GIPF), so we used these family descriptors to predict retention time of PCBs [40–43].

2. Materials and methods

2.1. Retention time of PCBs

The observed high resolution gas chromatographic relative retention times (HRGC-RRTs) of 209 PCBs on eighteen different stationary phases, 30m DB1, 30m SPB-Octyl, 60m SPB-Octyl, 100m CP-Sil5-C18, 30m DB5-MS, 60m RTX-5, 50m CP-Sil-13, 30m SPB-20, 30m HP-35, 60m RTX-35, 30m DB-17, 60m HP-1301, 30m DP-XLB, 30m DB-35-MS, 50m HT-8, 30m Apiezon L, 30m CNBP#2, 48m 007-23, reported by Frame [12,30] served as experimental data in this study and the HRGC-RRTs designed as a dependent variables. A complete list of the names and corresponding experimental RRT values of PCBs on each stationary phase has been categorized in Table S1 of the supplementary data.

2.2. Computer hardware and software

All calculations were run on an 2.5 GHz Intel® Core™2 Quad Q 8300 CPU with 2 GB of RAM using all four available cores under Windows XP operating system. The ISIS/Draw version 2.3 software was used for drawing the molecular structures [44]. Molecular modeling and geometry optimization were employed by HyperChem (version 7.1, HyperCube, Inc.) [45]. Gaussian98 program [46] was operated to optimize the molecular structure. The structures were optimized by the B3LYP method of density functional theory (DFT) at the level of 6-31G*. SPSS software (version 16.0, SPSS, Inc.) http://www.spss.com/ was used for elimination selection stepwise

MLR analysis and other calculations were performed in the MATLAB (version 7.0, Math Works, Inc.) environment.

2.3. Molecular descriptors generation and calculation

The selection of appropriate descriptors that are significantly related to the property of interest is very important for predictive QSRR models. The descriptors can be chosen using domain knowledge about the examined property, or the mathematical methods for the selection of descriptors can be applied. Dorofeeva et al. in 2005 demonstrated torsion angle between two phenyl rings in PCBs depend on existence chlorine atom in *ortho*-positions and chlorine atom in other positions have no significant effect on this angle. For PCBs with 1, 2, 3, and 4 chlorine in *ortho* position torsion angle between two phenyl rings in PCBs is near to 55.8–59.3°, 83.8–88.9°, 90.5–91.1° and 90.0–90.4°, respectively [47]. So we draw phenyl rings in PCBs with chlorine atoms in *ortho* position as orthogonal in HyperChem software and pre-optimized each molecule in it using AM1 method [45] semi-empirical theory level as prior step. Then we used the density functional theory (DFT) B3LYP/6-31G* functional/basis set combination as implemented in the Gaussian98 software package to re-optimize the molecular geometries of 209 PCBs and to compute the electrostatic potential $V(r)$ on the molecular surfaces defined by the 0.001 au contour of the electron density $\rho(r)$ [48]. Force constants and frequencies were considered to be sure that geometry is at minimum. The electrostatic potential $V_s(r)$ that is created in the space around a molecule by its nuclei and electrons is calculated by Eq. (1):

$$V_s(r) = \sum_A \frac{z_A}{|R_A - r|} - \int \frac{\rho(r')dr'}{|r' - r|} \tag{1}$$

where Z_A is the charge on nucleus A, located at R_A. The first term on the right side of Eq. (1) is the nuclear contribution to $V(r)$ and is positive, the second term is due to the electrons and is accordingly negative [41,48].

The average local ionization energy, $\bar{I}(r)$, is defined by Eq. (2):

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

$\rho_i(r)$ is the electronic density of the molecular orbital at the point r, ε_i is its orbital energy and $\rho(r)$ is the electronic density function.

We interpret $\bar{I}(r)$ as the energy required, on average, to remove an electron from a point r in the space of an atom or molecule [49,50]. $V_s(r)$ is effective for non-covalent interactions, which are largely electrostatic in nature, while $\bar{I}_s(r)$ is more suitable when there is transfer of charge (electron pair donor–electron pair acceptor interaction) that is one of forces responsible for separation of compounds in chromatography [32,49,51]. It might seem that $V_s(r)$ could also predict sites for electrophilic and nucleophilic bond-forming attack, by means of its most negative and positive regions. However $V_s(r)$ is not consistently reliable in this respect, because the regions of most negative $V_s(r)$ do not always correspond to the sites where the most reactive electrons are located. For example, the most negative $V_s(r)$ in benzene derivatives such as aniline, phenol, fluoro- and chlorobenzene, and nitrobenzene are associated with the substituents, whereas electrophilic reaction occurs on the rings. In contrast, $\bar{I}_s(r)$ correctly predicts the *ortho*/*para*- or *meta* directing effects of the substituents, as well as their activation or deactivation of the ring [49].

Politzer et al. demonstrated general interaction properties function (GIPF) can be summarized as the following equation [41,42]:

$$\text{Property} = f(V_{mv}, A_s^{tot}, A_s^+, A_s^-, V_{s,\max}, V_{s,\min}, \bar{V}_s, \bar{V}_s^+, \bar{V}_s^-, \pi^{tot}, \delta_{tot}^2, \delta_+^2, \delta_-^2, \nu, \bar{I}_{s,\max}, \bar{I}_{s,\min}, \bar{\bar{I}}_s, \delta_{\bar{I}_s}^2, \pi_{\bar{I}_s}) \tag{3}$$

In the present investigation, then we used the WFA statistical analysis program to compute GIPF descriptors using the produced CUBE file with Gaussian98 software package [49]. In Eq. (3), V_{mv} is the molecular volume and A_s^{tot}, A_s^+, A_s^- are total surface area and the surface area over which $V_s(r)$ is positive and negative, respectively. Politzer et al. showed molecular volume related to polarizability of molecule so they used this descriptor in GIPF approach [52,53]. $V_{s,max}$,$V_{s,min}$, respectively are the maxima and minima of electrostatic potential on the molecular surface and $\bar{V}_s$, $\bar{V}_s^+$ and $\bar{V}_s^-$, respectively are the overall average potentials and the average of positive and negative potentials and computed as:

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

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

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

π interpreted as an indicator of internal charge separation, which is present even in molecules having zero dipole moment due to symmetry, e.g. *para*-dinitrobenzene and boron trifluoride.

δ_{tot}^2, δ_+^2, and δ_-^2 are total, positive, and negative variances of electrostatic potentials, respectively and computed as:

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

ν is electrostatic balance parameter and computed as [53]:

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

In these summations, t is the total number of points on the surface grid, m and n are the numbers of points at which $V(r)$ is positive and negative, respectively [49]. The features of $\bar{I}(r)$ could be characterized analogously to those of $V(r)$ – its extrema $\bar{I}_{s,\max}$, $\bar{I}_{s,\min}$, its average magnitude $\bar{\bar{I}}_s$, average deviation ($\pi_{\bar{I}_s}$), and variance ($\delta_{\bar{I}_s}^2$) – keeping in mind that $\bar{I}(r)$ only has positive values [41,49,54,55].

3. Results and discussion

3.1. Selection descriptors and models developing

The calculated GIPF family descriptors are comprised of 15 surface electrostatic potential, 5 average local ionization energy and 16 combinations between them for a total of 36 descriptors, which were collected in a data matrix (**D**) with dimensions ($m \times n$), where m is the number of molecules and n is the number of descriptors. At the beginning, in order to minimize the information overlap in descriptors and to reduce the number of descriptors required in regression equation, the concept of non-redundant descriptors (NRD) [56] was used in our study. That is, when two descriptors are correlated by a linear correlation coefficient value >0.85, both descriptors are correlated with the dependent variables, the better correlation is used for the actual analysis, leaving out the descriptors showing a lower correlation. This objective-based feature selection left reduced and predictive descriptors for the studied compounds. By using these criteria, 21 out of 36 original descriptors were eliminated. These descriptors are not correlating with each other as revealed from the correlation matrix presented in Table S2 of the supplementary data. In GIPF approach, properties of molecule have relationship with few numbers of descriptors; therefore a variable reduction technique is needed. In this study the most important variables are selected by an elimination stepwise

selection procedure, which combines the forward selection and backward elimination approaches. This procedure considers first the descriptive variable most highly correlated with the response. If the inclusion of this variable results in a significant improvement of the regression model, evaluated with an overall *F*-test, it is retained and the selection continues. In a next step the variable that gives the largest significant decrease of the regression sum of squares, evaluated with a partial *F*-test, is added. After each forward selection step a backward elimination step is performed. In this step a partial *F*-test for the variables, already in the equation, is carried out. If a variable is no longer contributing significantly to the regression model, it is removed. The procedure stops at the moment that no variables fulfill the requirements anymore to be removed or entered. After this selection procedure classical MLR can be applied on the retained variables to build a predictive model [31,57]. An MLR model assumes that there is a linear relationship between the molecular descriptors of a compound, which is usually expressed as a feature vector x (with each descriptor as a component of this vector), and its target property, y. An MLR model can be described using the following equation:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 +, \cdots, +\beta_k X_k + \varepsilon \quad (8)$$

where $\{X_1, \ldots, X_k\}$ are molecular descriptors, β_0 is the regression model constant, β_1 to β_k are the coefficients corresponding to the descriptors X_1 to X_k and y is dependent variable. The values for β_0 to β_k are chosen by minimizing the sum of squares of the vertical distances of the points from the hyperplane so as to give the best prediction of y from X. Regression coefficients represent the independent contributions of each calculated molecular descriptor. In matrix notation, we will write the MLR model is defined in Eq. (9) as:

$$y = Xb + e \quad (9)$$

where $X_{(n\times k)}$ is of full column rank, including a column of 1s for the intercept if the intercept is included in the mean function y. We will further assume that we have selected a parameterization for the y so that X has full column rank, meaning that the inverse $(X^TX)^{-1}$ exists; this is not an important limitation on regression models because we can always delete terms from the y, or equivalently delete columns from X, until we have full rank. The $k \times 1$ vector b is the unknown parameter vector. The vector e consists of unobservable errors that we assume are equally variable and uncorrelated, unless stated otherwise. In appropriate Eq. (9), the least squares solution estimate b by $\hat{b} = (X^TX)^{-1}X^Ty$, and the fitted values y corresponding to the experimental RRTs are then given by:

$$\hat{y} = X\hat{b} = X(X^TX)^{-1}X^Ty = Hy \quad (10)$$

where H is the $n \times n$ matrix defined by $H = X(X^TX)^{-1}X^T$, which is called the hat matrix because it transforms the vector of experimental responses y into the vector of fitted responses $\hat{y}$. The vector of residuals $\hat{e}$ is defined by $\hat{e} = y - \hat{y} = y - X\hat{b} = y - X(X^TX)^{-1}X^Ty = (I - H)y$.

The advantages of MLR are that it is simple to use and the derived models are easy to interpret. The sign of the coefficients β_0 to β_k shows whether the molecular descriptors contribute positively or negatively to the target property and their magnitudes indicate the relative importance of the descriptors to the target property. However, the molecular descriptors should be mathematically independent (orthogonal) of one another and the number of compounds in the training set should exceed the number of molecular descriptors by at least a factor of 4 [58]. The reduced set of descriptors, remaining as consequence of previous steps implementation, was used to established best QSRR models by simple and multiple linear regressions, depending on the number of terms in model. After stepwise MLR calculation, we can model relative retention times of 209 PCBs with one descriptor that is common in all 18 column stationary phases. The resulted QSRR models of 209 PCBs with the exception of system 15 which contains 208 PCBs with selected four descriptors obtained for each stationary phase are given in Table 1. The value after the symbol "±" in the parenthesis is the standard deviation related to the regression coefficient. The qualities of the models derived from various subsets are evaluated using some statistics, such as the calibration squared correlation coefficients (R^2), root-mean-square error (RMSE), relative error of prediction (REP) and Fisher statistic ratio (F) are included in Table 1 for the best fitted equations. From Table 1, it can be seen, the predicted correlation coefficients (R^2) over 0.9819 with the exception of system 27 (R^2 = 0.8691) and the RMSE and REP below 0.0336 and 7.4217, respectively; except for system 27 (RMSE = 0.0494, REP = 10.7105) indicated that the best univariate linear regression models have good statistical qualities

Table 1
QSRR models and statistical parameters of GC-RRT values versus (52 + 180) for the total sets (n = 209) of PCB congeners on eighteen GC capillary columns.

St. ph.	Univariate models of the training sets	R^2	RMSE	REP	F	R^2_{CV}	$RMSE_{CV}$	R_{max}	$R_{CV,max}$
S1	RRT = −23.1934(±0.3368) + 2.1015(±0.0299) $\bar{I}_s$	0.9597	0.0304	6.6294	4930	0.9583	0.0309	0.0783	0.0599
S4	RRT = −20.5633(±0.2452) + 1.8683(±0.0218) $\bar{I}_s$	0.9726	0.0221	4.7743	7350	0.9719	0.0224	0.0378	0.0198
S6	RRT = −24.1530(±0.2815) + 2.1866(±0.0250) $\bar{I}_s$	0.9736	0.0254	5.5623	7640	0.9729	0.0257	0.0632	0.0439
S8	RRT = −23.6704(±0.2903) + 2.1438(±0.0258) $\bar{I}_s$	0.9709	0.0262	5.7269	6910	0.9702	0.0265	0.0752	0.0568
S10	RRT = −15.4163(±0.1502) + 1.4151(±0.0133) $\bar{I}_s$	0.9819	0.0135	2.6567	11,200	0.9816	0.0137	0.0713	0.0532
S11	RRT = −21.1968(±0.2254) + 1.9247(±0.0200) $\bar{I}_s$	0.9781	0.0203	4.3703	9230	0.9776	0.0206	0.0411	0.0231
S12	RRT = −21.6402(±0.2521) + 1.9641(±0.0224) $\bar{I}_s$	0.9738	0.0227	4.8880	7690	0.9732	0.0230	0.0650	0.0460
S13	RRT = −18.0583(±0.2028) + 1.6465(±0.0180) $\bar{I}_s$	0.9758	0.0183	3.8762	8350	0.9754	0.0185	0.0328	0.0149
S14	RRT = −22.3062(±0.3280) + 2.0232(±0.0291) $\bar{I}_s$	0.9588	0.0296	4.2389	4820	0.9577	0.0300	0.0266	0.0062
S15	RRT = −19.0572(±0.2297) + 1.7352(±0.0204) $\bar{I}_s$	0.9723	0.0207	4.3886	7230	0.9718	0.0209	0.0358	0.0167
S16	RRT = −14.3339(±0.1835) + 1.3162(±0.0163) $\bar{I}_s$	0.9692	0.0165	3.4508	6520	0.9687	0.0167	0.0307	0.0128
S17	RRT = −18.9243(±0.2421) + 1.7231(±0.0215) $\bar{I}_s$	0.9687	0.0218	4.6654	6420	0.968	0.0221	0.0503	0.0321
S20	RRT = −14.8553(±0.1617) + 1.3651(±0.0144) $\bar{I}_s$	0.9776	0.0146	2.8698	9020	0.9772	0.0147	0.0356	0.0167
S21	RRT = −13.6184(±0.1628) + 1.2525(±0.0145) $\bar{I}_s$	0.9731	0.0147	3.0709	7490	0.9727	0.0148	0.0312	0.0119
S22	RRT = −17.3659(±0.2174) + 1.5849(±0.0193) $\bar{I}_s$	0.9702	0.0196	4.1632	6730	0.9696	0.0198	0.0535	0.0354
S24	RRT = −24.8279(±0.3730) + 2.2463(±0.0331) $\bar{I}_s$	0.9569	0.0336	7.4217	4590	0.9558	0.0341	0.0460	0.0277
S26	RRT = −13.2136(±0.2363) + 1.2166(±0.0210) $\bar{I}_s$	0.9419	0.0213	4.4579	3360	0.9409	0.0215	0.0248	0.0076
S27	RRT = −19.8432(±0.5476) + 1.8041(±0.0487) $\bar{I}_s$	0.8691	0.0494	10.7105	1370	0.8665	0.0499	0.0334	0.0154

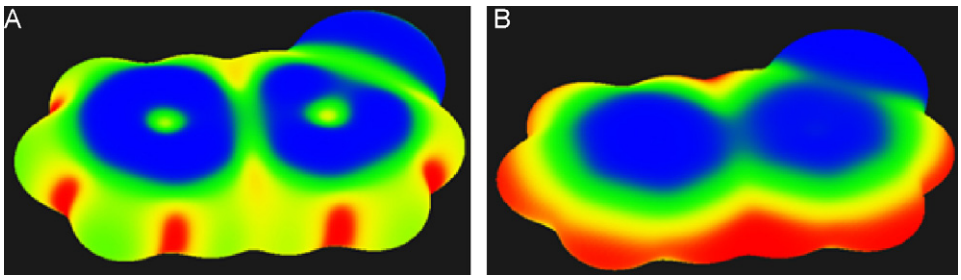

Fig. 2. (A) Calculated B3LYP/6-31G* ionization energy on molecular surface of 2biph. Ionization ranges in eV/mol: red, more than 12.9415; yellow, between 12.9415 and 11.6779; green, between 11.6779 and 10.4143; blue, smaller than 10.4143, (B) calculated B3LYP/6-31G* electrostatic potential molecular surface of 2biph. Electrostatic potential ranges in kcal/mol: red, more than 11.8356; yellow, between 11.8356 and 3.0794; green, between 3.0794 and −5.6769; blue, more negative than −5.6769. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

with low prediction error and demonstrated an excellent predictive power of the obtained QSRR models. This descriptor is average of average local ionization energy, $\bar{\bar{I}}_s$, also each local ionization energy descriptor (average deviation $\pi_{\bar{I}_s}$, variance ($\delta^2_{\bar{I}_s}$), minimum ($\bar{I}_{s,\min}$) of ionization energy) has high correlation ($R^2 \geq 0.79$) with this descriptor alone. As mentioned above local ionization energy is energy required, on average, to remove an electron from a point r in the space of an atom or molecule. This indicates interaction responsible for separation is charge transfer interaction that is because of high electron density on phenyl rings in PCBs (Figs. 2A and 3A) [41]. Charge transfer interaction (electron pair donor–electron pair acceptor interaction) is one of forces responsible for separation of compounds in chromatography. This interaction takes place, when a place exists with high electron density (Figs. 2B and 3B) (such as double bond or phenyl ring) and other with low electron density (positive charge) that withdraw electron cloud toward itself [39]. Examples of these interactions are C—H$\cdots\pi$, O—H$\cdots\pi$ interaction. For example C—H$\cdots\pi$ interaction was first proposed by Nishio and co-workers to explain the preference of conformations in which bulky alkyl and phenyl groups had close contact [59]. Statistical analysis of the crystal structure database indicates that the short contact of the C—H bond and the π system is observed in large numbers of organic crystals [61 and references therein]. In the preceding paper [38], we demonstrated retention time of PCBs depend on position and number of chlorine atom on phenyl ring in PCBs. There is high dependency between averages of molecular surface average local ionization and number and position of chlorine atoms in PCBs:

$$\bar{\bar{I}}_s = 11.0541(\pm 0.0018) + 0.0277(\pm 0.0005)\text{no. } o\text{-Cl} + 0.0463(\pm 0.0005)\text{no. } m\text{-Cl} + 0.0514(\pm 0.0008)\text{no. } p\text{-Cl}$$
$$R^2 = 0.9857,\ R^2_{CV} = 0.9850,\ \text{RMSE} = 0.0085,\ F = 4696.7 \quad (11)$$

where no. *o*-Cl, no. *m*-Cl and no. *p*-Cl are number of chlorine atoms on *ortho*, *meta* and *para* position, respectively. This research indicates $\bar{\bar{I}}_s$ can play role of these three descriptors in prediction retention time of PCBs.

3.2. Model prediction-validation

Model validation is a critical component of QSRR development. A number of procedures have been established to determine the quality of QSRR models. Therefore, a leave-one-out cross-validation (LOO-CV), *Y*-randomization, and external validation (EV) procedures through an odd–even number and division of the entire data set into training and test sets are used to validate the predictive ability and check the statistical significance of the developed 18 QSRR models.

3.2.1. Cross-validation

The robustness of the QSRR models can be tested by cross-validation procedures. These procedures eliminate one or several data sets (i.e. compounds) from the training set, derive a quantitative model from the remaining objects, and predict the relative retention time for the one or several objects which were not included in the derivation of the model [61,62]. The cross-validated squared correlation coefficient, R^2_{CV}, and the standard deviation of the predictions, S_{PRESS}, or root mean square error in cross-validation, $RMSE_{CV}$, are calculated from the predictive residual sum of squares, $PRESS = \sum (y_{pred} - y_{obs})^2$, in the same manner as R^2 and RMSE values are calculated from the unexplained variance $\sum (y_{cal} - y_{obs})^2$, to describe the quality of fit of the models. The formula used to calculate the R^2_{CV} and $RMSE_{CV}$, is given below:

$$R^2_{CV} = 1 - \frac{\sum_{i=1}^{N}(y_{pred,i} - y_{obs,i})^2}{\sum_{i=1}^{N}(y_{obs,i} - \bar{y}_{obs})^2} = 1 - \frac{PRESS}{\sum_{i=1}^{N}(y_{obs,i} - \bar{y}_{obs})^2} \quad (12)$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{N}(y_{pred,i} - y_{obs,i})^2}{N}} \quad (13)$$

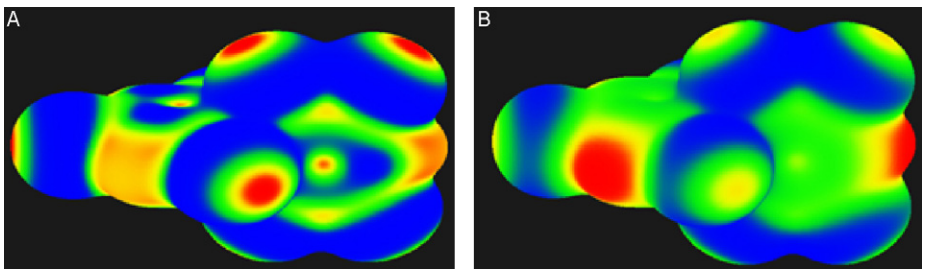

Fig. 3. (A) Calculated B3LYP/6-31G* ionization energy on molecular surface of 188biph. Ionization ranges in eV/mol: red, more than 13.0502; yellow, between 13.0502 and 12.0948; green, between 12.0948 and 11.1394; blue, smaller than 11.1394, (B) calculated B3LYP/6-31G* electrostatic potential molecular surface of 188biph. Electrostatic potential ranges in kcal/mol: red, more than 11.8356; yellow, between 11.8356 and 3.0794; green, between 3.0794 and −5.6769; blue, more negative than −5.6769. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

where N is the number of training patters, $y_{obs,i}$ and $y_{pred,i}$ are the experimental, and predicted RRTs of the left-out PCB i, respectively and $\bar{y}_{obs}$ is the average experimental RRT of left-in PCBs different from i. In the leave-one-out procedure, only one object is eliminated at a time and the process is repeated until all objects have been eliminated once and only once. For larger data sets, the elimination of several objects at a time, randomly or in a systematic manner, is recommended. Throughout this investigation, only leave-one-out cross-validation was performed to derive a measure of the internal predictivity of the models, within the training set. Values of R^2_{CV} can range from 1 to <0. A value of one indicates a perfect prediction, and a value of 0 means that the QSRR derived has no modeling power. Negative values arise from a situation where the derived QSRR is a poorer description of data than no model at all. The R^2_{CV} values can be considered as a measure of the predictive power of a model: whereas R^2 can always be increased artificially by adding more parameters, R^2_{CV} decreases if a model is over parameterized [63], and is therefore a more meaningful summary statistic for predictive models. The correlation coefficients (R^2_{CV}) and $RMSE_{CV}$ for each subset are presented in Table 1. The cross-validation results show that the R^2_{CV} are higher than 0.8665 and $RMSE_{CV}$ lower than 0.0499 for all GC stationary phases, respectively. Furthermore, in all cases, the cross-validated R^2_{CV} values are very close to the corresponding R^2 values and the cross-validated $RMSE_{CV}$ values are only slightly larger than the corresponding RMSE values. Clearly, the cross-validation demonstrates the final models to be statistically significant and this considered as an ultimate proof for the high predictive power of QSRR models [58].

3.2.2. Y-randomization test

The Y-randomization of response is another important validation approach that is widely used to establish model robustness [64]. As a much more reliable criterion for the risk of chance correlations, the affinity values of the PCBs were re-ordered in a random manner (y scrambling), to determine the percentage of chance correlations that are as good as or even better than the best correlations found for the y values in their correct order; 100 different randomization runs were performed routinely, and the best models were individually derived for each randomization by systematic search, including up to three columns of the $N \times N$ matrices in the regression models. In some cases, even 1000 randomizations were performed to prove that 100 randomizations give representative results. If all QSRR models obtained in the Y-randomization test have relatively high R_{max} and $R_{CV,max}$, it implies that an acceptable QSRR model cannot be obtained for the given data set by current modeling method. No attempt was made to eliminate y-scrambled data sets with (fortuitous) high correlations between the real, original y values and the randomized y values. The results are shown in last two column in Table 1. Very low level of R_{max} (in the interval of 0.0248 for S26 and 0.0783 for S1) and $R_{CV,max}$ (in the interval of 0.0062 for S14 and 0.0599 for S1) indicates good results in our original models and is not due to a chance correlation or structural dependency of the training set for each stationary phase of GC column [31].

3.2.3. Odd–even external validation

To validate and develop a believable QSRR model, it is not enough to build a model for the whole data set. Consequently, the 209 (208 for system 15) data set for all stationary phases were sorted in the ascending order of RRT values and then divided into two sets namely “odd set” and “even set” RRTs [65,66]. This way of splitting ensures that the distribution of RRT values of the two subsets were very similar. The QSRR models were fitted to the odd set and even set samples separately and the resulted fitness was assessed by applying QSRR models to both samples. To compare the estimation abilities of the models, two statistical parameters namely root mean square error (RMSE) and R^2, were calculated. The same data set (i.e. ‘calibration set’) that was already used to fit the models was employed to determine resubstitution parameters, i.e. $RMSE_{RS}$ and R^2_{RS}, also to determine holdout parameters, i.e. $RMSE_{HO}$ and R^2_{HO} for the other data set, which was not involved in the fitting. The resubstitution statistical parameters of the samples base their predictions on the regression fitted to those samples and this is while the holdout statistical parameters base their predictions on the regression fitted to the other samples. The plots of RRTs estimated by odd- and even-set QSRR models (holdout prediction) versus the RRTs observed experimentally are given in Fig. 4, also Table 2 summarizes these statistical parameters achieved by this approach. As can be seen, in the odd and even-set samples, the resubstitution and holdout RMSE are very similar, indicating that the same sample and other sample predictions are equally precise for all stationary phases [37,38].

3.2.4. Selecting training and test sets

We should emphasize that all three cross-validation, Y-randomization and odd–even external validation must be made a mandatory part of model development. This goal can be completed by a division of an experimental RRT data set into the training and

Table 2
Statistical parameters of the over-fitting and predictive ability of the models.

St. ph.	Odd samples				Even samples			
	$RMSE_{RS}$	R^2_{RS}	$RMSE_{HO}$	R^2_{HO}	$RMSE_{RS}$	R^2_{RS}	$RMSE_{HO}$	R^2_{HO}
S1	0.0332	0.9539	0.0332	0.9539	0.0274	0.9662	0.0274	0.9662
S4	0.0244	0.9677	0.0244	0.9677	0.0196	0.9779	0.0197	0.9779
S6	0.0278	0.9693	0.0279	0.9693	0.0228	0.9783	0.0228	0.9783
S8	0.0287	0.9662	0.0287	0.9662	0.0235	0.9760	0.0235	0.9760
S10	0.0141	0.9810	0.0142	0.9810	0.0129	0.9832	0.0130	0.9832
S11	0.0221	0.9750	0.0221	0.9750	0.0184	0.9816	0.0184	0.9816
S12	0.0244	0.9709	0.0244	0.9709	0.0209	0.9772	0.0210	0.9772
S13	0.0192	0.9742	0.0192	0.9742	0.0173	0.9778	0.0173	0.9778
S14	0.0318	0.9542	0.0318	0.9542	0.0272	0.9641	0.0273	0.9641
S15	0.0206	0.9724	0.0206	0.9724	0.0209	0.9722	0.0209	0.9722
S16	0.0163	0.971	0.0164	0.9710	0.0166	0.9681	0.0167	0.9681
S17	0.0236	0.9647	0.0237	0.9647	0.0199	0.9733	0.0199	0.9733
S20	0.0153	0.9760	0.0154	0.9760	0.0138	0.9794	0.0138	0.9794
S21	0.0149	0.9733	0.0149	0.9733	0.0145	0.9732	0.0145	0.9732
S22	0.0209	0.9673	0.0209	0.9673	0.0183	0.9734	0.0183	0.9734
S24	0.0363	0.9514	0.0364	0.9514	0.0308	0.9628	0.0309	0.9628
S26	0.0224	0.9378	0.0224	0.9378	0.0203	0.9463	0.0203	0.9463
S27	0.0514	0.8632	0.0515	0.8632	0.0475	0.8755	0.0475	0.8755

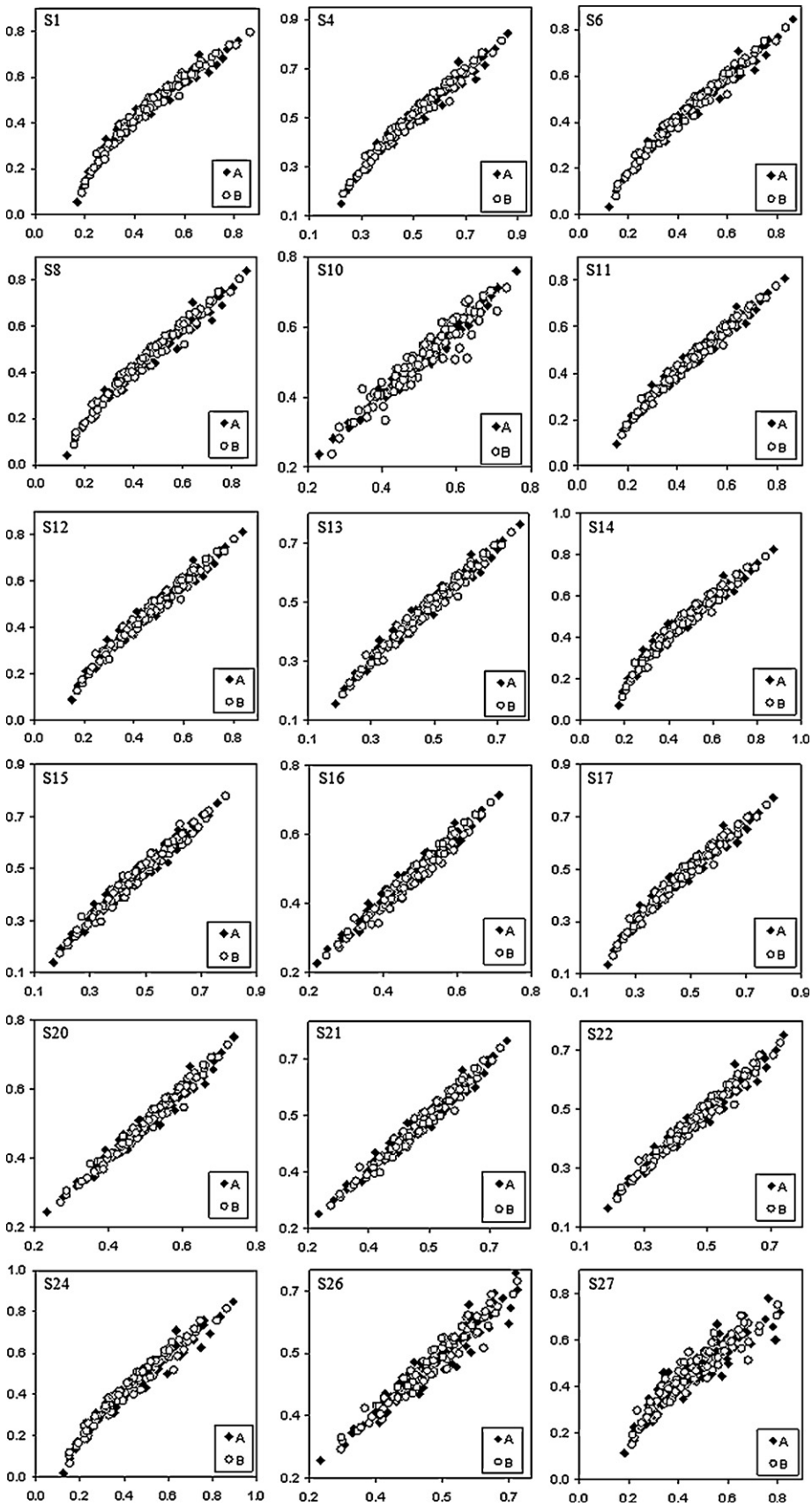

Fig. 4. Plots of the RRTs estimated for the odd set (♦, A) and even set (○, B) samples by holdout model versus that observed RRTs experimentally for all stationary phases of 209 PCBs.

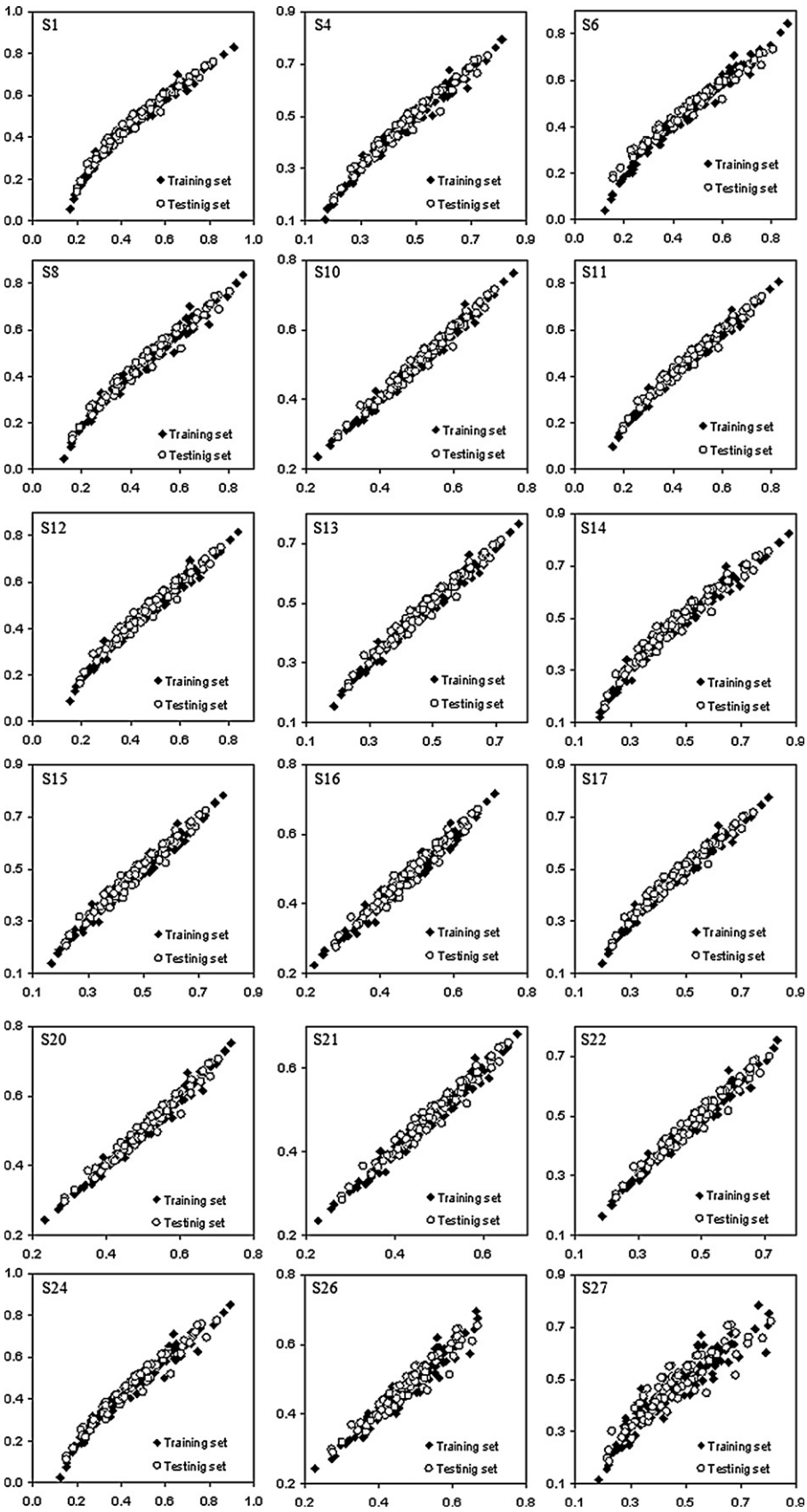

Fig. 5. Plots of the RRTs estimated by the QSPR models in Table 3 versus that observed for 126 training set PCBs (♦) and 83 (system 15 82) testing set ones (○) for all stationary phases.

Table 3
QSRR models and statistical parameters of GC-RRT values versus (52 + 180) for the training (n = 126) and testing sites of PCB congeners on eighteen GC capillary columns.

St. ph.	Training set							Testing set		
	Univariate models	RMSE	REP	F	R^2	R^2_{CV}	$RMSE_{CV}$	RMSE	REP	R^2
S1	$RRT = -22.9159(\pm0.4319) + 2.0771(\pm0.0384)\,\bar{\bar{I}}_s$	0.0310	6.8375	2928.0	0.9594	0.9569	0.0320	0.0296	6.3730	0.9625
S4	$RRT = -20.3907(\pm0.3123) + 1.8531(\pm0.0278)\,\bar{\bar{I}}_s$	0.0224	4.8909	4457.3	0.9729	0.9718	0.0229	0.0217	4.6240	0.9732
S6	$RRT = -23.9089(\pm0.3630) + 2.1651(\pm0.0323)\,\bar{\bar{I}}_s$	0.0261	5.7834	4504.7	0.9732	0.9720	0.0267	0.0359	7.7316	0.9755
S8	$RRT = -23.4171(\pm0.3732) + 2.1215(\pm0.0332)\,\bar{\bar{I}}_s$	0.02689	5.9342	4090.8	0.9706	0.9693	0.0274	0.0253	5.4470	0.9726
S10	$RRT = -15.4247(\pm0.1801) + 1.4160(\pm0.0160)\,\bar{\bar{I}}_s$	0.0129	2.5551	7827.5	0.9844	0.9839	0.0131	0.0145	2.8129	0.9781
S11	$RRT = -21.0793(\pm0.2812) + 1.9145(\pm0.0250)\,\bar{\bar{I}}_s$	0.0202	4.3852	5868.1	0.9793	0.9784	0.0206	0.0207	4.3915	0.9771
S12	$RRT = -21.5370(\pm0.3117) + 1.9552(\pm0.0277)\,\bar{\bar{I}}_s$	0.0224	4.8580	4982.4	0.9757	0.9747	0.0229	0.0235	4.9751	0.9718
S13	$RRT = -18.0208(\pm0.2465) + 1.6433(\pm0.0219)\,\bar{\bar{I}}_s$	0.0177	3.7817	5626.3	0.9784	0.9777	0.0181	0.0193	4.0410	0.9723
S14	$RRT = -22.1330(\pm0.4113) + 2.0081(\pm0.0366)\,\bar{\bar{I}}_s$	0.0296	6.4250	3018.1	0.9605	0.9586	0.0303	0.0299	6.3613	0.9579
S15	$RRT = -19.0372(\pm0.2808) + 1.7337(\pm0.0250)\,\bar{\bar{I}}_s$	0.0202	4.3081	4824.9	0.9749	0.9741	0.0205	0.0217	4.5488	0.9691
S16	$RRT = -14.3848(\pm0.2232) + 1.3209(\pm0.0198)\,\bar{\bar{I}}_s$	0.0160	3.3641	4435.2	0.9728	0.9720	0.0163	0.0174	3.5945	0.9639
S17	$RRT = -18.7949(\pm0.3013) + 1.7118(\pm0.0268)\,\bar{\bar{I}}_s$	0.0216	4.6608	4086.8	0.9706	0.9693	0.0221	0.0223	4.7073	0.9672
S20	$RRT = -14.8588(\pm0.1947) + 1.3655(\pm0.0173)\,\bar{\bar{I}}_s$	0.0140	2.7712	6229.7	0.9805	0.9799	0.0142	0.0155	3.0124	0.9728
S21	$RRT = -13.6561(\pm0.1922) + 1.2560(\pm0.0171)\,\bar{\bar{I}}_s$	0.0138	2.9049	5406.1	0.9776	0.9769	0.0140	0.0160	3.1396	0.9661
S22	$RRT = -17.3065(\pm0.2650) + 1.5797(\pm0.0236)\,\bar{\bar{I}}_s$	0.0191	4.0789	4497.0	0.9732	0.9723	0.0194	0.0205	4.2902	0.9656
S24	$RRT = -24.5666(\pm0.4836) + 2.2233(\pm0.0430)\,\bar{\bar{I}}_s$	0.0348	7.7595	2675.3	0.9557	0.9537	0.0355	0.0320	6.9229	0.9600
S26	$RRT = -13.2018(\pm0.2874) + 1.2156(\pm0.0255)\,\bar{\bar{I}}_s$	0.0207	4.3480	2264.2	0.9481	0.9464	0.0210	0.0222	4.6023	0.9314
S27	$RRT = -19.7384(\pm0.6604) + 1.7951(\pm0.0587)\,\bar{\bar{I}}_s$	0.0475	10.3561	935.4	0.8830	0.8789	0.0483	0.0522	11.2029	0.8482

For system 15 n = 125.

test sets, which are used for model development and validation, respectively. In this investigation, for further testing the predictive ability of the models for the external compounds without the models, part of the congeners are picked up from 209 (208 for system 15) PCBs to construct a training set which is used to develop a prediction model and then predict the values of RRTs in the remaining congeners. How to pick up the compounds in the training set is very important for developing of the predictive QSRR models. In this case, before each training run, all data sets were split randomly into two separate sub-matrices: the training set matrix and external testing set matrix. Out of 126 congeners (60.3%) were used for the training set and 83 congeners (82 for system 15) (39.7%) were used as external validation. The PCBs constituting the training and testing sets are clearly presented in Table S1. Moreover, the same divisions were repeated with corresponding RRTs values. The test examples are marked as bold font and training set was also used to obtain the best fit equation of simple linear regression with average of average local ionization energy, $\bar{\bar{I}}_s$ descriptor. Furthermore, the testing set was used to monitor overfitting the models. The resulted models for training set congeners were the same as those obtained for the entire set of all PCBs in each subset subject to use descriptors of all congener's models supporting sufficient ability for the prediction set of 83 PCBs. The resulting regression equations of the training set for individual HRGC column stationary phases are indexed in Table 3, and the results are plotted in Fig. 5. Statistical parameters for the best-fitted models are also presented in Table 3. The correlation coefficients (R^2) of the obtained models are >0.95 for all the stationary phases except for system 27 (0.8830), and the highest one is 0.9844 for system 10. The root mean square error (RMSE) and relative error prediction (REP) of estimation ranged from 0.0129, 2.5551 of system 10 to 0.0348, 7.7595 of system 24 (except for system 27), respectively, also the F statistic values are >2264.2 (except system 27). The LOO-CV method was used to examine the stability of QSRR models, and the values of R^2_{CV} and $RMSE_{CV}$ for the models were above 0.9464 and in the range of 0.0131 and 0.0355 (except for system 27).

The predicted RRTs versus the observed RRTs of the 126 PCB training sets are plotted in Fig. 5 (diamond). As shown in Table 3 and Fig. 5, the QSRR statistical results exhibit good estimation capacity and stability for internal training set PCB samples to individual stationary phases. High predictive ability of QSRR models for external examples is another criterion of a good QSRR model. The predicted RRTs of 83 (82 for system 15) PCBs in the external testing set by the models in Table 3 are also demonstrated in Fig. 5 (circle) versus the observed RRTs of 18 GC stationary phases. For all 18 HRGC stationary phases, the regression of the observed and predicted RRTs had a high agreement with the diagonal of each chart. The predicted correlation coefficients (R^2) over 0.9314 with the exception of system 27 (R^2 = 0.8482), the root mean square error (RMSE) and relative error prediction (REP) below 0.0359 and 7.7316, respectively, except for system 27 (RMSE = 0.0522, REP = 11.2029) demonstrated an excellent predictive power of the obtained QSRR models.

4. Conclusions

Based on the GIPF family descriptors derived from properties of the electrostatic potentials on their molecular surfaces computed at the B3LYP/6-31G* level, QSRR models were built to study the HRGC-RRT behaviors of all 209 PCBs congeners on 18 capillary stationary phases (S1, S4, S6, S8, S10, S11, S12, S13, S14, S15, S16, S17, S20, S21, S22, S24, S26, S27), by simple linear equations. As has just been discussed, we have found good relationships between HRGC-RRT values and our computed the average of average molecular surface local ionization energy ($\bar{\bar{I}}_s$) value, for each stationary phases taken separately. The presence of $\bar{\bar{I}}_s$ term in all QSRR models may indicate the charge-transfer interactions effect is responsible for separation in all of these 18 capillary stationary phases. The validation and predictive ability of the models were examined by three methods of leave-one-out cross-validation, Y-randomization, and external validation. The methods indicated that the resulted simple linear equation QSRR models have high prediction ability and low overfitting. The results demonstrate system 10 has best performance for separation of PCBs because have maximum R^2 and minimum RMSE although these statistical parameters have no significance different with these statistics parameter for other stationary phase (except for system 27). System 27 has the weakest performance for PCBs separation, although complete separation of all PCBs are not

possible and recommended this system don't used for separation of PCBs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2012.01.047.

References

[1] IPCS, Environmental Health Criteria No. 140. Polychlorinated Biphenyls, International Programme on Chemical Safety, 1992, http://www.inchem.org/documents/ehc/ehc/ehc48.htm.
[2] HSDB, Hazardous substances data bank. National Library of Medicine. http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?, 2009. HSDB and search on CAS number (accessed 25.08.09).
[3] E.M. Silberhorn, H.P. Glauert, L.W. Robertson, Crit. Rev. Toxicol. 20 (1990) 440.
[4] D.O. Carpenter, Rev. Environ. Health 21 (2006) 1.
[5] U.S. Department of Health and Human Services Secretary Kathleen Sebelius, 12th Report on Carcinogens (RoC), 2011, 349. http://ntp.niehs.nih.gov/go/roc12.
[6] S.H. Safe, Crit. Rev. Toxicol. 24 (1994) 87.
[7] S. Knerr, D. Schrenk, Crit. Rev. Toxicol. 36 (2006) 663.
[8] M.S. Denison, S.R. Nagy, Annu. Rev. Pharmacol. Toxicol. 43 (2003) 309.
[9] J. Matthews, B. Wihlén, N. Heldring, L. Macpherson, L. Helguero, E. Treuter, L.A. Haldosén, J.Å. Gustafsson, Biochem. J. (Printed in Great Britain) 406 (2007) 343.
[10] WHO, Polychlorinated Biphenyls and Terphenyls. Environmental Health Criteria, vol. 140, 2nd ed., World Health Organization, Geneva, 1993.
[11] W. Vetter, B. Luckas, J. Buijten, J. Chromatogr. A 799 (1998) 249.
[12] G.M. Frame, Fresenius J. Anal. Chem. 357 (1997) 714.
[13] R.L. Grob, E.F. Barry, Modern Practice of Gas Chromatography, 4th ed., Wiley-Interscience, 2004.
[14] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987.
[15] J.W. McFarland, C.M. Berger, S.A. Froshauer, S.F. Hayashi, S.J. Hecker, B.H. Jaynes, M.R. Jefson, B.J. Kamicker, C.A. Lipinski, K.M. Lundy, C.P. Reese, C.B. Vu, J. Med. Chem. 40 (1997) 1340.
[16] P. Gramatica, N. Navas, R. Todeschini, Chemometr. Intell. Lab. Syst. 40 (1998) 53.
[17] R. Kaliszan, M.A. van Straten, M. Markuszewski, C.A. Cramers, H.A. Claessens, J. Chromatogr. A 855 (1999) 455.
[18] T. Hyötyläinen, M. Kallio, K. Hartonen, M. Jussila, S. Palonen, M.-L. Riekkola, Anal. Chem. 74 (2002) 4441.
[19] B. Larsen, S. Bowadt, R. Tilio, in: J. Albaigés (Ed.), Environmental Analytical Chemistry of PCBs, vol. 16, Gordon and Breach Sci. Pub., Singapore, 1993, p. 3.
[20] E. Storr-Hansen, in: J. Albaigés (Ed.), Environmental Analytical Chemistry of PCBs, vol. 16, Gordon and Breach Sci. Pub., Singapore, 1993, p. 24.
[21] R. Eganhouse, B. Gould, D. Olaguer, C. Phinney, P. Sherblom, in: J. Albaigés (Ed.), Environmental Analytical Chemistry of PCBs, vol. 16, Gordon and Breach Sci. Pub., Singapore, 1993, p. 111.
[22] B. Pavoni, A. Sfriso, S. Raccanelli, in: J. Albaigés (Ed.), Environmental Analytical Chemistry of PCBs, vol. 16, Gordon and Breach Sci. Pub., Singapore, 1993, p. 101.
[23] D. Bedard, R. May, Environ. Sci. Technol. 30 (1996) 237.
[24] J. Plomley, M. Lauševic, R. March, Mass Spectrom. Rev. 19 (2000) 305.
[25] R. March, J. Mass Spectrom. 32 (1997) 351.
[26] Z. Liu, J.B. Philips, J. Chromatogr. Sci. 29 (1991) 227.
[27] J.B. Phillips, J. Xu, J. Chromatogr. A 703 (1995) 327.
[28] P. Haglund, M. Harju, R. Ong, P. Marriott, J. Microcolumn Sep. 13 (2001) 306.
[29] J.F. Focant, A. Sjödin, D.G. Patterson, J. Chromatogr. A 1040 (2004) 227.
[30] G.M. Frame, Fresenius J. Anal. Chem. 357 (1997) 701.
[31] M. Makino, Chemosphere 39 (1999) 893.
[32] M.N. Hasan, P.C. Jurs, Anal. Chem. 60 (1988) 978.
[33] S.S. Liu, Y. Liu, D.Q. Yin, X.D. Wang, L.S. Wang, J. Sep. Sci. 29 (2006) 296.
[34] A. Krawczuk, A. Voelkel, J. Lulek, R. Urbaniak, K. Szyrwinska, J. Chromatogr. A 1018 (2003) 63.
[35] Y. Ren, H. Liu, X. Yao, M. Liu, Anal. Bioanal. Chem. 388 (2007) 165.
[36] L. Jäntschi, S.D. Bolboacă, M.V. Diudea, Int. J. Mol. Sci. 8 (2007) 1125.
[37] R. Ghavami, F. Sadeghi, Chromatographia 70 (2009) 851.
[38] R. Ghavami, S.M. Sajadi, Chromatographia 72 (2010) 523.
[39] R. Kaliszan, Chem. Rev. 107 (2007) 3212.
[40] P. Politzer, J.S. Murray, Theor. Chem. Acc. 108 (2002) 134.
[41] P. Politzer, J.S. Murray, F.A. Bulat, J. Mol. Model. 16 (2010) 1731.
[42] P. Politzer, J.S. Murray, Fluid Phase Equilibria 185 (2001) 129.
[43] P. Politzer, Y. Ma, P. Lane, M.C. Concha, Int. J. Quantum Chem. 105 (2005) 341.
[44] ISIS Draw 2.3, MDL Information Systems, Inc., 1990–2000.
[45] HyperChem Release 7.1 for Windows Molecular Modeling System Program Package, HyperCube, 2002.
[46] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dappich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. Petersson, P.Y. Aayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rubuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A.5, Gaussian, Inc., Pittsburgh, PA, 1998.
[47] O.V. Dorofeeva, V.P. Novikov, N.F. Moiseeva, V.S. Yungman, J. Struct. Chem. 46 (2005) 237.
[48] J.S. Murray, P. Politzer, M.C. Concha, J. Mol. Model. 13 (2007) 643.
[49] F.A. Bulat, A. Toro-Labbé, T. Brinck, J.S. Murray, P. Politzer, J. Mol. Model. 16 (2010) 1679.
[50] P. Kulshrestha, N. Sukumar, J.S. Murray, R.F. Giese, T.D. Wood, J. Phys. Chem. A 113 (2009) 756.
[51] Y. Ma, K.C. Gross, C.A. Hollingsworth, P.G. Seybold, J.S. Murray, J. Mol. Model. 10 (2004) 235.
[52] P. Jin, J.S. Murray, P. Politzer, Int. J. Quantum Chem. 96 (2004) 394.
[53] O.G. Gonzalez, J.S. Murray, Z. Peralta-Inga, P. Politzer, Int. J. Quantum Chem. 83 (2001) 115.
[54] P. Politzer, J.S. Murray, F. Abu-Awwad, Int. J. Quantum Chem. 76 (2000) 643.
[55] P.J.T. Brinck, J.S. Murray, P. Politzer, Int. J. Quantum Chem. 95 (2003) 632.
[56] J. Olivero, T. Garcia, P. Payares, R. Vivas, D. Diaz, E. Daza, P. Geerliger, J. Pharm. Sci. 86 (1997) 625.
[57] M. Dumarey, A.M. van Nederkassel, E. Deconinck, Y. Vander Heyden, J. Chromatogr. A 1192 (2008) 81.
[58] R.P. Verma, C. Hansch, Eur. J. Med. Chem. 45 (2010) 1470.
[59] Y. Kodama, K. Nishihata, M. Nishio, N. Nakagawa, Tetrahedron Lett. (1977) 2105.
[60] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, J. Am. Chem. Soc. 122 (2000) 3746.
[61] H. Kubinyi, F.A. Hamprecht, T. Mietzner, J. Med. Chem. 41 (1998) 2553.
[62] M. Clark, R.D. Cramer, Quant. Struct. —Act. Relat. 12 (1993) 137.
[63] D.M. Hawkins, S.C. Basak, D. Mills, J. Chem. Inf. Comput. Sci. 43 (2003) 579.
[64] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22 (2003) 69.
[65] D.M. Hawkins, J. Chem. Inf. Comput. Sci. 44 (2004) 1.
[66] I.V. Tetko, D.J. Livingstone, A.I. Luik, J. Chem. Inf. Comput. Sci. 35 (1995) 826.