CMST 22(1) 41-53 (2016) DOI:10.12921/cmst.2016.22.01.004

Computational Model For Chromatographic Relative Retention Time of Polychlorinated Biphenyls Using Sub-structural Molecular Fragments

Saadi Saaidpour

Department of Chemistry, Faculty of Science
Sanandaj Branch, Islamic Azad University, Sanandaj, Iran
**E-mail: sasaaidpour@iausdj.ac.ir*

Received: 10 February 2016; revised: 17 March 2016; accepted: 17 March 2016; published online: 24 March 2016

Abstract: Quantitative structure-retention relationship (QSRR) analysis is a useful technique capable of relating chromatographic retention time to the chemical structure of a solute. Using the sub-structural molecular fragments (SMF) derived directly from the molecular structures, the gas chromatographic relative retention times (RRTs) of 209 polychlorinated biphenyls (PCBs) on the SE-54 stationary phase were calculated. An eight-variable regression equation with the correlation coefficient of 0.9945 and the root mean square errors of 0.0134 was developed. Forward and backward stepwise regression variable selection and multi-linear regression analysis (MLRA) are combined to describe the effect of molecular structure on the RRT of PCB according to the QSRR method. To quantitatively relate RRT with the molecular structure MLR analysis is performed on the set of 163 sub-structural molecular fragments (SMF) provided by the ISIDA software. The eight fragments selected by variable subset selection, all belonging to the sub-fragments, adequately represent the structural factors influencing the affinity of PCB to SE-54 stationary phase in the separation process. Finally, a QSRR model is selected based on leave-one-out cross-validation and its prediction ability is further tested on 42 representative compounds excluded from model calibration. The prediction results from the MLR model are in good agreement with the experimental values. By applying the MLR method we can predict the test set with squared cross validated correlation coefficient (Q^2_{ext}) of 0.9913 and root mean square error (RMSE) of 0.0169.

Key words: polychlorinated biphenyl, relative retention time, sub-structural molecular fragment, MLR, QSRR

I. INTRODUCTION

Polychlorinated biphenyls (PCB) that were widely used in industry as dielectric fluids in transformers and capacitors, as hydraulic and heat transfer fluids, and as plasticizers, are now of concern as prevalent, persistent, and toxic pollutants [1, 2]. PCB produced widespread global contamination of water and soil and bio-accumulated in food chains due to their high hydrophobicity and chemical stability [3, 4]. The following examples emphasize the importance of bioaccumulation of PCB and their high impact on different aspects of the biosphere pollution. Contamination of surface soil by PCB remains a serious problem in Dalian, Liaoning Province, China [5]. Not only are surface soils exposed to PCB contamination, but various aquatic species are endangered as well used isotope dilution HRGC/HRMS method to determine polybrominated/chlorinated biphenyls (Co-PXB) in 18 different Japanese fish fillets [6]. The eggs of San Francisco Bay aquatic birds contain high PCB concentrations [7]. PCB congeners are adversely affecting sediments and the crab population density in mangroves near Rio de Janeiro, Brazil [8]. PCB concentration can be used to assess risks related to the exposure to other persistent bio-accumulative and toxic compounds. Mori et al. [9] showed that the total PCB concentration in human blood is potentially a reliable indicator of the total dioxin concentration, which is of special concern in Japan because of the Kanemi Yusho tragedy [10]. Polychlorinated biphenyls (PCBs) are a class of discrete or-

ganic compounds with one to ten chlorine atoms attached to a biphenyl nucleus and a general chemical formula of $C_{12}H_{10-n}Cl_n$, where $n = 1 \div 10$ [11]. A general chemical structure of polychlorinated biphenyls is shown in Fig. 1.

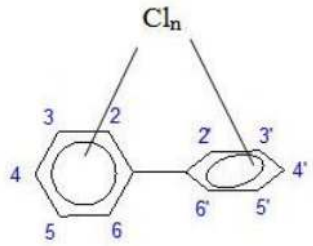

Fig. 1. General structural formula and substitution positions of the PCBs

The composition of PCBs is summarized in Tab. 1 [12]. PCBs are hydrophobic compounds with low volatility, and the highly chlorinated ones have poor water solubility. Moreover, they are resistant to acids, bases, and (generally) environmental degradation processes. They are, therefore, highly persistent in the environment. A series of properties and activities of PCBs have been investigated by QSPR/QSAR modeling: aqueous solubility [13], gas/particle partitioning in the atmosphere [14], photo degradation half-life in n-hexane solution under UV irradiation [15], n-octanol/water partition coefficients [16, 17], vaporization [18, 19], and sublimation enthalpy [20]. The retention time of PCB congeners has also been previously investigated and reported [21-24]. Due to the need to control the PCBs level in the environment, one of the most commonly used methods for their analysis in environmental samples is gas chromatography coupled with an electron-capture detector, because of its high sensitivity toward halogenated compounds [25, 26], but easy identification of individual congeners remains unresolved for the moment [27,28]. 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. Based on this approach, many authors have described multiple regression models for predicting gas-chromatographic relative retention time [RRT] on a SE-54 stationary phase using different kinds of molecular descriptors [29-31]. A range of empirical and semi-empirical tools have been developed for the prediction of retention behavior of different classes of compounds under various chromatographic conditions. Many of these predictive models fall into the category of quantitative structure-retention relationships (QSRR) which derive relationships between chromatographic parameters and molecular structure properties (descriptors) of the analytes. These quantitative structure property relationships (QSPR) are generally used to correlate the biological, chemical, or physical property of a compound with its physico-chemical characteristics. In some of our previous papers, we reported on the application of QSPR techniques to develop a new, simplified approach to prediction of compound properties [32-41]. For the first time we applied the sub-structural molecular fragment (SMF) method for modeling gas chromatographic relative retention times of PCBs. The goal of this study is to develop an SMF method and the related software tools to model relationships between the structure of 209 polychlorinated biphenyls and their relative retention times on the SE-54 stationary phase. This method is based on to represent a molecule by its fragments and on to calculate their contributions to a given property. It uses two types of fragments: (i) the sequences of atoms and/or bonds (atom and/or bond paths up to specified maximal length) and (ii) "augmented" represented by a selected atom and/or bonds with its environment. In fact, it represents an extension of empirical methods used to calculate physical or chemical properties of molecules using atomic or bond increments.

II. DATA AND METHODS

To undertake QSRR studies two kinds of input data are needed. One is a set of quantitatively comparable retention data (dependent variable) for a sufficiently large (for the statistical reason) set of analytes. The other is a set of quantities (independent variables) assumed to account for structural differences among the studied analytes. Through the use of chemometric computational techniques, retention parameters are characterized in terms of various descriptors of analytes (and/or their combinations) or in terms of systematic knowledge extracted (learnt) from these descriptors. To obtain statistically significant and physically meaningful QSRR, reliable input data are required and a stringent mathematical analysis must be carried out. The great advantage of the QSRR analysis over other quantitative structure property relationship studies is that chromatography can readily produce a large amount of relatively precise and reproducible data. In a chromatographic process all conditions may be kept constant and hence the structure of an analyte becomes the single independent variable in the system.

The QSRR model for the estimation of the RRT of PCB congeners is established in the following six steps: the molecular structure input and generation of the files containing the chemical structures is stored in a computer-readable format; quantum mechanics geometry is optimized with a semi-empirical (AM1) method; sub-structural molecular fragments are computed; molecular fragments are selected; and the molecular fragments – RRT model is generated by the multi-linear regression analysis (MLRA), and statistical approval techniques and prediction analysis.

II. 1. Data set

The relative response times of all PCBs obtained by using temperature-programmed, high resolution gas chromatography on a capillary column of SE-54 (Methyl 5% Phenyl

Tab. 1. Composition of PCBs by homologs

Homolog	Molecular formula	Chlorine(%by weight)	Number of isomers
Monochlorobiphenyl	$C_{12}H_9Cl$	19	3
Dichlorobiphenyl	$C_{12}H_8Cl_2$	32	12
Trichlorobiphenyl	$C_{12}H_7Cl_3$	41	24
Tetrachlorobiphenyl	$C_{12}H_6Cl_4$	49	42
Pentachlorobiphenyl	$C_{12}H_5Cl_5$	54	46
Hexachlorobiphenyl	$C_{12}H_4Cl_6$	59	42
Heptachlorobiphenyl	$C_{12}H_3Cl_7$	63	24
Octachlorobiphenyl	$C_{12}H_2Cl_8$	66	12
Nonachlorobiphenyl	$C_{12}HCl_9$	69	3
Decachlorobiphenyl	$C_{12}Cl_{10}$	71	1
Total congeners	-	-	209

poly-siloxane, non-polar), reported by Mullin et al. [29,42] served as experimental data in this study. The values were used as a dependent variable in the following analyses and the values ranged from 0.1544 to 1.0496. The names of the compounds used in this study with their relative retention times are listed in Tab. 2.

Tab. 2. Experimental data of relative retention times of polychlorinated biphenyls

No	Molecule	RRT(exp)	No	Molecule	RRT(exp)
1	2-chloro-1,1'-biphenyl	0.1544	2	3-chloro-1,1'-biphenyl	0.1937
3	4-chloro-1,1'-biphenyl	0.1975	4	2,6-dichloro-1,1'-biphenyl	0.2243
5	2,2'-dichloro-1,1'-biphenyl	0.2245	6	2,4-dichloro-1,1'-biphenyl	0.2566
7	2,5-dichloro-1,1'-biphenyl	0.257	8	2,3'-dichloro-1,1'-biphenyl	0.2709
9	2,4'-dichloro-1,1'-biphenyl	0.2783	10	2,3-dichloro-1,1'-biphenyl	0.2785
11	3,5-dichloro-1,1'-biphenyl	0.2973	12	2,6,2'-trichloro-1,1'-biphenyl	0.3045
13	2,4,6-trichloro-1,1'-biphenyl	0.3165	14	3,3'-dichloro-1,1'-biphenyl	0.3238
15	3,4-dichloro-1,1'-biphenyl	0.3298	16	3,4'-dichloro-1,1'-biphenyl	0.3315
17	2,5,2'-trichloro-1,1'-biphenyl	0.3378	18	4,4'-dichloro-1,1'-biphenyl	0.3387
19	2,4,2'-trichloro-1,1'-biphenyl	0.3398	20	2,3,6-trichloro-1,1'-biphenyl	0.3508
21	2,6,3'-trichloro-1,1'-biphenyl	0.3521	22	2,3,2'-trichloro-1,1'-biphenyl	0.3625
23	2,6,4'-trichloro-1,1'-biphenyl	0.3636	24	2,3,5-trichloro-1,1'-biphenyl	0.377
25	2,3',5'-trichloro-1,1'-biphenyl	0.3782	26	2,6,2',6'-tetrachloro-1,1'-biphenyl	0.38
27	2,4,5-trichloro-1,1'-biphenyl	0.382	28	2,5,3'-trichloro-1,1'-biphenyl	0.3911
29	2,4,3'-trichloro-1,1'-biphenyl	0.3937	30	2,4,6,2'-tetrachloro-1,1'-biphenyl	0.4007
31	2,5,4'-trichloro-1,1'-biphenyl	0.4024	32	2,4,4'-trichloro-1,1'-biphenyl	0.4031
33	2,3,4-trichloro-1,1'-biphenyl	0.4135	34	2,3',4'-trichloro-1,1'-biphenyl	0.4163
35	2,3,3'-trichloro-1,1'-biphenyl	0.417	36	2,5,2',6'-tetrachloro-1,1'-biphenyl	0.4187
37	2,4,2',6'-tetrachloro-1,1'-biphenyl	0.4242	38	2,3,4'-trichloro-1,1'-biphenyl	0.4267
39	2,3,6,2'-tetrachloro-1,1'-biphenyl	0.4334	40	3,5,3'-trichloro-1,1'-biphenyl	0.4375
41	2,3,2',6'-tetrachloro-1,1'-biphenyl	0.445	42	3,5,4'-trichloro-1,1'-biphenyl	0.4488
43	2,4,6,3'-tetrachloro-1,1'-biphenyl	0.451	44	2,6,3',5'-tetrachloro-1,1'-biphenyl	0.4554
45	2,5,2',5'-tetrachloro-1,1'-biphenyl	0.4557	46	2,3,5,2'-tetrachloro-1,1'-biphenyl	0.4587
47	3,4,5-trichloro-1,1'-biphenyl	0.4593	48	2,4,2',5'-tetrachloro-1,1'-biphenyl	0.461
49	2,4,2',4'-tetrachloro-1,1'-biphenyl	0.4639	50	2,4,6,4'-tetrachloro-1,1'-biphenyl	0.4643
51	2,4,5,2'-tetrachloro-1,1'-biphenyl	0.4651	52	2,3,5,6-tetrachloro-1,1'-biphenyl	0.4671
53	2,3,4,6-tetrachloro-1,1'-biphenyl	0.4685	54	3,4,3'-trichloro-1,1'-biphenyl	0.4738
55	2,4,6,2',6'-pentachloro-1,1'-biphenyl	0.4757	56	2,3,2',5'-tetrachloro-1,1'-biphenyl	0.4832
57	3,4,4'-trichloro-1,1'-biphenyl	0.4858	58	2,3,6,3'-tetrachloro-1,1'-biphenyl	0.486
59	2,3,2',4'-tetrachloro-1,1'-biphenyl	0.487	60	2,5,3',5'-tetrachloro-1,1'-biphenyl	0.4984
61	2,6,3',4'-tetrachloro-1,1'-biphenyl	0.4989	62	2,3,4,2'-tetrachloro-1,1'-biphenyl	0.499
63	2,3,6,4'-tetrachloro-1,1'-biphenyl	0.4999	64	2,4,3',5'-tetrachloro-1,1'-biphenyl	0.504
65	2,3,6,2',6'-pentachloro-1,1'-biphenyl	0.5057	66	2,3,2',3'-tetrachloro-1,1'-biphenyl	0.5102

Tab. 2 – continued:

No	Molecule	RRT(exp)	No	Molecule	RRT(exp)
67	2,4,6,2',5'-pentachloro-1,1'-biphenyl	0.5142	68	2,3,5,3'-tetrachloro-1,1'-biphenyl	0.5155
69	2,4,6,2',4'-pentachloro-1,1'-biphenyl	0.5212	70	2,4,5,3'-tetrachloro-1,1'-biphenyl	0.5214
71	2,3,3',5'-tetrachloro-1,1'-biphenyl	0.5267	72	2,3,5,4'-tetrachloro-1,1'-biphenyl	0.529
73	2,3,4,5-tetrachloro-1,1'-biphenyl	0.5331	74	2,3,5,2',6'-pentachloro-1,1'-biphenyl	0.5331
75	2,4,5,4'-tetrachloro-1,1'-biphenyl	0.5341	76	2,5,3',4'-tetrachloro-1,1'-biphenyl	0.5407
77	2,3',4',5'-tetrachloro-1,1'-biphenyl	0.5408	78	2,3,2',4',6'-pentachloro-1,1'-biphenyl	0.5415
79	2,4,5,2',6'-pentachloro-1,1'-biphenyl	0.5431	80	2,3,5,6,2'-pentachloro-1,1'-biphenyl	0.5437
81	2,4,3',4'-tetrachloro-1,1'-biphenyl	0.5447	82	3,5,3',5'-tetrachloro-1,1'-biphenyl	0.5464
83	2,3,6,2',5'-pentachloro-1,1'-biphenyl	0.5464	84	2,3,4,6,2'-pentachloro-1,1'-biphenyl	0.5486
85	2,4,6,3',5'-pentachloro-1,1'-biphenyl	0.5518	86	2,3,6,2',4'-pentachloro-1,1'-biphenyl	0.5549
87	2,3,4,3'-tetrachloro-1,1'-biphenyl	0.5562	88	2,4,6,2',4',6'-hexachloro-1,1'-biphenyl	0.5666
89	2,3,3',4'-tetrachloro-1,1'-biphenyl	0.5676	90	2,3,4,4'-tetrachloro-1,1'-biphenyl	0.5676
91	2,3,5,2',5'-pentachloro-1,1'-biphenyl	0.5742	92	2,3,6,2',3'-pentachloro-1,1'-biphenyl	0.5744
93	2,3,4,2',6'-pentachloro-1,1'-biphenyl	0.5779	94	2,3,5,2',4'-pentachloro-1,1'-biphenyl	0.5814
95	2,4,5,2',5'-pentachloro-1,1'-biphenyl	0.5816	96	2,3,6,3',5'-pentachloro-1,1'-biphenyl	0.5862
97	2,4,5,2',4'-pentachloro-1,1'-biphenyl	0.588	98	3,4,3',5'-tetrachloro-1,1'-biphenyl	0.5894
99	2,4,6,3',4'-pentachloro-1,1'-biphenyl	0.5968	100	2,3,6,2',4',6'-hexachloro-1,1'-biphenyl	0.5969
101	2,3,5,6,3'-pentachloro-1,1'-biphenyl	0.5986	102	2,3,4,6,3'-pentachloro-1,1'-biphenyl	0.6016
103	3,4,5,3'-tetrachloro-1,1'-biphenyl	0.6024	104	2,3,5,2',3'-pentachloro-1,1'-biphenyl	0.6029
105	2,3,5,6,2',6'-hexachloro-1,1'-biphenyl	0.6062	106	2,3,2',4',5'-pentachloro-1,1'-biphenyl	0.61
107	2,3,4,5,2'-pentachloro-1,1'-biphenyl	0.6105	108	2,3,4,5,6-pentachloro-1,1'-biphenyl	0.6132
109	2,6,3',4',5'-pentachloro-1,1'-biphenyl	0.6142	110	3,4,5,4'-tetrachloro-1,1'-biphenyl	0.6149
111	2,3,4,6,2',6'-hexachloro-1,1'-biphenyl	0.6149	112	2,3,5,6,4'-pentachloro-1,1'-biphenyl	0.615
113	2,3,4,6,4'-pentachloro-1,1'-biphenyl	0.6171	114	2,3,4,2',5'-pentachloro-1,1'-biphenyl	0.6175
115	2,3,5,3',5'-pentachloro-1,1'-biphenyl	0.6183	116	2,3,4,2',4'-pentachloro-1,1'-biphenyl	0.6224
117	2,3,5,2',4',6'-hexachloro-1,1'-biphenyl	0.6243	118	2,4,5,3',5'-pentachloro-1,1'-biphenyl	0.6256
119	2,3,6,2',3',6'-hexachloro-1,1'-biphenyl	0.6257	120	3,4,3',4'-tetrachloro-1,1'-biphenyl	0.6295
121	2,3,6,3',4'-pentachloro-1,1'-biphenyl	0.6314	122	2,4,5,2',4',6'-hexachloro-1,1'-biphenyl	0.6349
123	2,3,4,2',3'-pentachloro-1,1'-biphenyl	0.6453	124	2,3,5,6,2',5'-hexachloro-1,1'-biphenyl	0.6499
125	2,3,5,2',3',6'-hexachloro-1,1'-biphenyl	0.6563	126	2,3,4,6,2',5'-hexachloro-1,1'-biphenyl	0.6563
127	2,5,3',4',5'-pentachloro-1,1'-biphenyl	0.6584	128	2,3,5,6,2',4'-hexachloro-1,1'-biphenyl	0.6608
129	2,3,4,3',5'-pentachloro-1,1'-biphenyl	0.6626	130	2,3,5,3',4'-pentachloro-1,1'-biphenyl	0.6628
131	2,4,3',4',5'-pentachloro-1,1'-biphenyl	0.6658	132	2,3,6,2',4',5'-hexachloro-1,1'-biphenyl	0.6672
133	2,3,4,5,3'-pentachloro-1,1'-biphenyl	0.668	134	2,4,5,3',4'-pentachloro-1,1'-biphenyl	0.6693
135	2,3,4,6,2',4'-hexachloro-1,1'-biphenyl	0.6707	136	2,3,4,2',4',6'-hexachloro-1,1'-biphenyl	0.6707
137	2,3,4,5,2',6'-hexachloro-1,1'-biphenyl	0.6789	138	2,3,5,6,2',3'-hexachloro-1,1'-biphenyl	0.6796
139	2,3,4,5,4'-pentachloro-1,1'-biphenyl	0.6828	140	2,3,4,5,6,2'-hexachloro-1,1'-biphenyl	0.6848
141	2,3,4,6,2',3'-hexachloro-1,1'-biphenyl	0.6853	142	2,3,3',4',5'-pentachloro-1,1'-biphenyl	0.6871
143	2,3,5,2',3',5'-hexachloro-1,1'-biphenyl	0.6871	144	2,3,5,6,3',5'-hexachloro-1,1'-biphenyl	0.692
145	2,3,5,6,2',4',6'-heptachloro-1,1'-biphenyl	0.692	146	2,3,5,2',4',5'-hexachloro-1,1'-biphenyl	0.6955
147	2,3,4,6,3',5'-hexachloro-1,1'-biphenyl	0.6968	148	2,3,4,6,2',4',6'-heptachloro-1,1'-biphenyl	0.7016
149	2,3,4,2',3',6'-hexachloro-1,1'-biphenyl	0.7035	150	2,4,5,2',4',5'-hexachloro-1,1'-biphenyl	0.7036
151	2,3,4,3',4'-pentachloro-1,1'-biphenyl	0.7049	152	2,4,6,3',4',5'-hexachloro-1,1'-biphenyl	0.7068
153	3,4,5,3',5'-pentachloro-1,1'-biphenyl	0.7078	154	2,3,4,5,2',5'-hexachloro-1,1'-biphenyl	0.7203
155	2,3,5,6,2',3',6'-heptachloro-1,1'-biphenyl	0.7205	156	2,3,4,2',3',5'-hexachloro-1,1'-biphenyl	0.7284
157	2,3,4,6,2',3',6'-heptachloro-1,1'-biphenyl	0.7305	158	2,3,4,5,2',4'-hexachloro-1,1'-biphenyl	0.7329
159	2,3,4,5,6,3'-hexachloro-1,1'-biphenyl	0.7396	160	2,3,5,6,3',4'-hexachloro-1,1'-biphenyl	0.7396
161	2,3,6,3',4',5'-hexachloro-1,1'-biphenyl	0.7399	162	2,3,4,2',4',5'-hexachloro-1,1'-biphenyl	0.7403
163	2,3,4,5,6,2',6'-heptachloro-1,1'-biphenyl	0.7416	164	2,3,4,6,3',4'-hexachloro-1,1'-biphenyl	0.7429

Tab. 2 – continued:

No	Molecule	RRT(exp)	No	Molecule	RRT(exp)
165	2,3,4,5,2',3'-hexachloro-1,1'-biphenyl	0.7501	166	3,4,5,3',4'-pentachloro-1,1'-biphenyl	0.7512
167	2,3,5,6,2',3',5'-heptachloro-1,1'-biphenyl	0.7537	168	2,3,4,5,6,4'-hexachloro-1,1'-biphenyl	0.7572
169	2,3,4,6,2',3',5'-heptachloro-1,1'-biphenyl	0.7611	170	2,3,4,5,2',4',6'-heptachloro-1,1'-biphenyl	0.7653
171	2,3,5,6,2',4',5'-heptachloro-1,1'-biphenyl	0.7654	172	2,3,4,5,3',5'-hexachloro-1,1'-biphenyl	0.7655
173	2,3,4,6,2',4',5'-heptachloro-1,1'-biphenyl	0.772	174	2,3,5,3',4',5'-hexachloro-1,1'-biphenyl	0.7737
175	2,3,4,2',3',4'-hexachloro-1,1'-biphenyl	0.7761	176	2,4,5,3',4',5'-hexachloro-1,1'-biphenyl	0.7814
177	2,3,4,5,6,2',5'-heptachloro-1,1'-biphenyl	0.7848	178	2,3,4,5,2',3',6'-heptachloro-1,1'-biphenyl	0.7965
179	2,3,4,5,6,2',4'-heptachloro-1,1'-biphenyl	0.7968	180	2,3,4,2',3',5',6'-heptachloro-1,1'-biphenyl	0.8031
181	2,3,4,6,2',3',4'-heptachloro-1,1'-biphenyl	0.8089	182	2,3,5,6,2',3',5',6'-octachloro-1,1'-biphenyl	0.8089
183	2,3,4,5,3',4'-hexachloro-1,1'-biphenyl	0.8105	184	2,3,4,5,6,2',3'-heptachloro-1,1'-biphenyl	0.8152
185	2,3,4,3',4',5'-hexachloro-1,1'-biphenyl	0.8184	186	2,3,4,6,2',3',5',6'-octachloro-1,1'-biphenyl	0.8197
187	2,3,4,5,6,2',4',6'-octachloro-1,1'-biphenyl	0.8217	188	2,3,4,5,6,3',5'-heptachloro-1,1'-biphenyl	0.8269
189	2,3,4,5,2',3',5'-heptachloro-1,1'-biphenyl	0.8278	190	2,3,4,6,2',3',4',6'-octachloro-1,1'-biphenyl	0.8293
191	2,3,4,5,2',4',5'-heptachloro-1,1'-biphenyl	0.8362	192	2,3,5,6,3',4',5'-heptachloro-1,1'-biphenyl	0.8397
193	2,3,4,6,3',4',5'-heptachloro-1,1'-biphenyl	0.8447	194	2,3,4,5,6,2',3',6'-octachloro-1,1'-biphenyl	0.8494
195	3,4,5,3',4',5'-hexachloro-1,1'-biphenyl	0.8625	196	2,3,4,5,2',3',4'-heptachloro-1,1'-biphenyl	0.874
197	2,3,4,5,6,3',4'-heptachloro-1,1'-biphenyl	0.874	198	2,3,4,5,6,2',3',5'-octachloro-1,1'-biphenyl	0.8845
199	2,3,4,5,2',3',5',6'-octachloro-1,1'-biphenyl	0.8875	200	2,3,4,5,2',3',4',6'-octachloro-1,1'-biphenyl	0.8938
201	2,3,4,5,6,2',4',5'-octachloro-1,1'-biphenyl	0.8938	202	2,3,4,5,3',4',5'-heptachloro-1,1'-biphenyl	0.9142
203	2,3,4,5,6,2',3',5',6'-nonachloro-1,1'-biphenyl	0.932	204	2,3,4,5,6,2',3',4'-octachloro-1,1'-biphenyl	0.9321
205	2,3,4,5,6,2',3',4',6'-nonachloro-1,1'-biphenyl	0.9423	206	2,3,4,5,2',3',4',5'-octachloro-1,1'-biphenyl	0.962
207	2,3,4,5,6,3',4',5'-octachloro-1,1'-biphenyl	0.9678	208	2,3,4,5,6,2',3',4',5'-nonachloro-1,1'-biphenyl	1.0103
209	2,3,4,5,6,2',3',4',5',6'-decachloro-1,1'-biphenyl	1.0496			

II. 2. Computer Hardware and Software

All calculations were run on a Dell Inspiron N5010 laptop computer with Intel® Core^{TM} i7 processor with Windows 7 operating system.

II. 3. Computational Procedure

II. 3. 1. Sub-structural Molecular Fragments

The ISIDA/QSPR program realizes the sub-structural molecular fragments (SMF) method [43-49] which is based on the splitting of a molecular graph on fragments (sub-graphs) and on the calculation of their contributions to a given property Y. Two classes of fragments are used: "sequences" (**I**) and "augmented" (**II**). Three sub-types **AB**, **A** and **B** are defined for each class. For fragments **I**, they represent sequences of atoms and bonds (**AB**), of atoms only (**A**), or of bonds only (**B**). The shortest or all paths from one atom to the other are used. For each type of sequences, the minimal (n_{min}) and maximal (n_{max}) number of constituted atoms must be defined. Thus, for the partitioning **I** (**AB**,

$n_{min} - n_{max}$), **I** (**A**, n_{min} - n_{max}) and **I** (**B**, n_{min} - n_{max}), the program generates "intermediate" sequences involving n atoms ($n_{min} \leq n \leq n_{max}$). In the current version of ISIDA/QSPR, $n_{min} \geq 2$ and $n_{max} \leq 15$. An "augmented" represents a selected atom with its environment including both neighboring atoms and bonds (AB), or atoms only (A, without taking hybridization of neighbors into account, or Hy, where hybridization of neighbors is accounted for), or bonds only (B).

II. 3. 2. Variable Selection Procedures

The generated pool of descriptors is generally much larger than the number of compounds in the training set; therefore procedures for selecting variables should be applied to build statistically significant multi-linear regressions. In ISIDA, a combination of forward and backward stepwise variable selection procedures is used.

1). *Filtering stage.* The program eliminates variables X_i which have small correlation coefficient with the property, $R_{y,i} < R^0_{y,i}$, and those highly correlated with other variables $X_j P_{i,j} > R^0_{i,j}$), which were already selected for the model. In this work, the values $R^0_{y,i} = 0.001, \ldots$ and $R^0_{i,j} = 0.75, \ldots$ were used. Fragments always occurring in the same combination in each compound of the training set (concatenated fragments) are treated as one extended fragment.

2). *Forward stepwise pre-selection stage*. The suite of forward and backward stepwise algorithms has been used for variable pre-selection in ISIDA/QSPR studies by the variable selection suite (VSS) program. Three algorithms for forward stepwise variable selection are based on the calculations of correlation coefficients and subtractions. This is an iterative procedure, on each step of which the program selects one X_i (two X_i and X_j or three variables X_i, X_j and X_k) maximizing the correlation coefficient $R_{y,j}$ ($R_{y,ij}$ or $R_{y,ijk}$) between X_i (X_i and X_j or X_i , X_j and X_k) and dependent variable Y. At the first step ($s = 1$), the modeled property for each compound is taken as its experimental one $Y_s = Y$. At each next step s, as the property value Y_s is used residual $Y_s = Y_{s-1} - Y_{\mathrm{calc}}$, where $Y_{\mathrm{calc}} = c_i X_i$($Y_{\mathrm{calc}} = c_i X_i + c_j X_j$ or $Y_{\mathrm{calc}} = c_i X_i + c_j X_j + c_k X_k$). is the calculated property by the one-variable (two- or three-variables) model with selected variable X_i (variables X_i and X_j or X_i, X_j and X_k). This loop is repeated until the number of variables k reaches a user-defined value; in this work, k was analyzed from $0.1n$ to $0.9n$, where n is the number of the molecules in the training set.

3). *Backward stepwise selection stage.* The final selection is performed using backward stepwise variable selection procedure based on the t statistic criterion. Here, the program eliminates the variables with low $t_i = c_i/s_i$ values, where s_i is standard deviation for the coefficient c_i at the i-th variable in the model. First, the program selects the variable with the smallest $t < t_0$, then it performs a new fitting excluding that variable. This procedure is repeated until $t \geq t_0$ for selected variables or if the number of variables reaches the user's defined value. Here, t_0, the tabulated value of the Student's criterion is a function of the number of data points, the number of variables, and the significance level. Default value of the t_0 is 1.96; it can be analyzed from 1.96 to 3.9.

II. 3. 3. Multi-linear Regression Model

The modeled physical or chemical property *Y* can be quantitatively calculated accounting for contributions of fragments using linear (1) fitting equation.

$$Y = \sum_i A_i \times N_i, \quad \text{Additive Model}, \tag{1}$$

where A_i is a fragment a contribution, N_i is the number of fragments of *i* type. Contributions of A_i are calculated by minimizing a functional

$$U[a_i] = \sum_{i=1}^{n} w_i \left(Y_{\mathrm{exp},i} - Y_{\mathrm{calc},i}\right)^2 => \min, \tag{2}$$

where *n* is the number of compounds in the training set, w_i the weight accounting for the accuracy of the experimental data, Y_{exp} and Y_{calc} are, respectively, experimental and calculated according to property values. The equation (1) represents calculation of property *Y* by using additive contributions of fragments. The coefficients of the equation (1) being optimized at the training stage are then used to estimate *Y* values of the compounds from the test set or to screen external databases of real or virtual compounds.

Using the singular value decomposition method (SVD), ISIDA/QSPR fits the a_i terms in equation (2) and calculates corresponding statistical characteristics (correlation coefficient (R), standard deviation (s), Fischer's criterion (F), cross-validation correlation coefficient (Q), standard deviation of predictions (S_{press}), Kubyni's criterion (*FIT*), R_H-factor of Hamilton and matrix of pair correlations (covariation matrix) for the terms a_i) and performs statistical tests to select the best model. The prediction ability of the model is characterized by leave-one-out correlation coefficient Q^2 and by leave-one-out standard deviation ($SDEP$), as well as by dispersions of predicted values of $\langle Y_{\mathrm{pred}} \rangle$ averaged over several models.

II. 3. 4. Validation of QSRR Model

In ISIDA/QSRR calculations, each initial data set was split into two sub-sets: training (167 compounds) and test (42 compounds) sets. The QSRR models were built on the training set followed by "prediction" calculations for the test set. Before a QSRR model is used to predict the properties for new compounds, it should be validated both internally and externally to ensure that the built model is robust, reliable, stable and predictive. In the current work, several statistic terms

Tab. 3. Set of fragments, Coefficients (Ai) of the equation, standard deviations for coefficients and their t-Test for $\mathrm{RRT} = \sum(A_i \times N_i)$

No	Variable[i]	Contribution (Ai)	Standard deviation (ΔA)	t-Test
1	C-C=C-H	0.0050	0.0005	10.63
2	Cl-C=C-H	0.0587	0.0025	23.58
3	Cl-C-C-H	0.0635	0.0024	26.06
4	Cl-C-C=C-C-H	0.0133	0.0012	11.37
5	Cl-C-C=C-H	0.0171	0.0013	13.03
6	Cl-C-C-Cl	0.1380	0.0044	31.57
7	Cl-C-C-C=C-H	0.0184	0.0030	6.06
8	Cl-C=C-Cl	0.1274	0.0044	29.07

such as squared correlation coefficient R^2 for the training set fitness and $Q^2{}_{\mathrm{ex}}$ for the external predictive ability, leave-one-out (LOO) cross-validated $Q^2{}_{\mathrm{LOO}}$ and root mean square error (RMSE) were used to assess the internal and external predictive ability of the proposed model. The corresponding statistical parameters were defined as:

$$R^2 = 1 - \frac{\sum_i^n \left(y_{ip} - y_{ie}\right)^2}{\sum_i^n \left(y_{ie} - y_{\mathrm{mean}}^{\mathrm{training}}\right)^2} \tag{3}$$

$$Q^2_{\mathrm{LOO}} = 1 - \frac{\sum_i^n \left(y_{ip} - y_{icv}\right)^2}{\sum_i^n \left(y_{ie} - y_{\mathrm{mean}}^{\mathrm{training}}\right)^2} (4) \tag{4}$$

$$Q^2_{\mathrm{ext}} = 1 - \frac{\sum_i^n \left(y_{ip} - y_{ie}\right)^2}{\sum_i^n \left(y_{ie} - y_{mean}^{test}\right)^2} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i^n \left(y_{ip} - y_{ie}\right)^2}{n}}, \tag{6}$$

where i represents i^{th} molecule, y_{ie} is the desired output (experimental property), y_{ip} the actual output, y_{icv} is the output of leave-one-out cross-validation, $y_{\mathrm{training}}^{\mathrm{mean}}$ and $y_{\mathrm{test}}^{\mathrm{mean}}$ are the mean values of y_{ip} for the training and test sets, respectively. N is the number of compounds in the training or test set. In addition, the built model was also validated externally using the test set compounds due to the fact that the best way to evaluate the predictive ability of a QSRR model is its validation using compounds not included in the training set with known properties.

III. RESULTS AND DISCUSSION

The ISIDA program has been developed to establish structure-retention relationship based on the SMF partitioning. The program inputs data in the SDF format [50] containing structural and properties information. The graphical interface of ISIDA allows to attribute data to the learning or validation sets, and to set up the parameters of calculations (type of fragments, minimal and maximal number of atoms/bonds in the sequences, type of equation). A QSRR is a mathematical relationship between a property of a chemical, in this

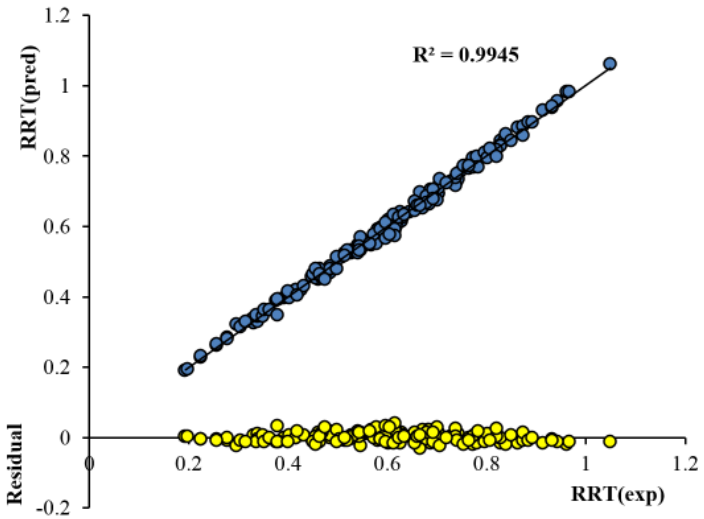

Fig. 2. Plot of predicted RRT and residuals versus experimental RRT of training set

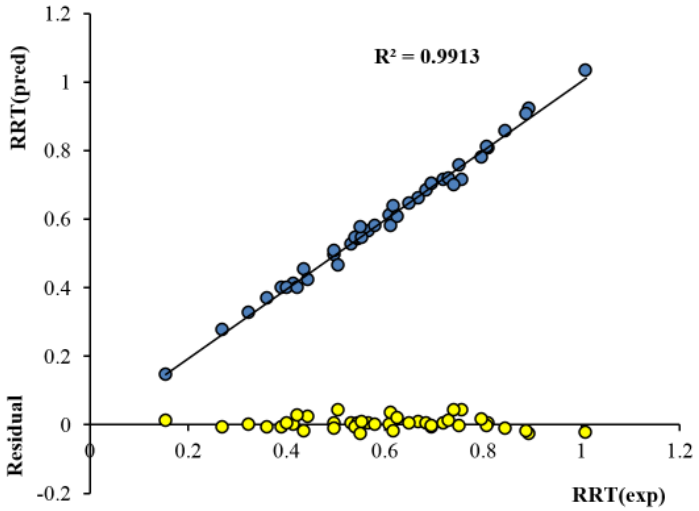

Fig. 3. Plot of predicted RRT and residuals versus experimental RRT of test set

case relative retention times, and molecular fragments of the chemical. The fragments are obtained from the structure of the chemical. First, a training set of 167 compounds is used to statistically establish the relationship between RRT and the molecular fragments. The QSRR can then be used to predict the retention times of test set (42 compounds) for which the fragments are known. Thus the fragments selected to describe this process in a QSRR should be able to describe the relative affinities of chemicals for the stationary phase. To establish relationships between the structure of PCBs and their retention times, we used the recently developed substructural molecular fragments (SMF) method which is based on the representation of the molecular graph by fragments and on the calculation of their contributions to a given property. The sequences fragments represent sequences of atoms and bonds (AB), of atoms only (A), or of bonds only (B). The length of sequences varies from 2 to 15 atoms. For any sequence containing from $n_{\min}$ to $n_{\max}$ atoms, all fragments of $n_{\max}$, $n_{\max-1}$, $n_{\max-2}$, ..., $n_{\min}$ length are considered. In this work, the **I** (**AB**, **4-6**) decomposition scheme corresponds to eight sequences containing 4, 5 and 6 atoms and linking bonds are selected. To select the most relevant fragments to the RRTs, 15 groups of fragments calculated by ISIDA for each compound were used as the inputs for stepwise regression. The optimum subset size was reached when adding another fragment did not significantly improve performance of the model. Through this procedure, the 8-parameter model was selected as the best model. It can be described in Tab. 2. The quality of a QSRR model is generally expressed by its fitting ability and prediction ability, and the latter one is more important. Statistical parameters for the test set were ${Q^2}_{ext}$ of 0.9913 and the standard deviation error of prediction (SDEP) of 0.0139. When a compound is split into constitutive fragments, the fragments contributions to the RRT or to any other physical or chemical property are calculated using linear fitting equation:

$$\mathrm{RRT} = \sum (A_i \times N_i). \tag{7}$$

Here, A_i is contribution of fragment, and N_i is the number of fragments of i type. The fragments contributions as fitted coefficients in the equation (9) at the learning stage are used to predict RRT for the compounds from the validation set. The set of fragments, coefficients of the equation, standard deviations for coefficients and their t-test for equation (9) are shown in Tab. 3.

The experimental, predicted and residuals data for the training set (167 compounds) and the test set (42 compounds) are shown in Tab. 4 and 5. The plot of predicted RRT and residuals versus experimental RRT of the training set and the test set are showed in Fig. 2 and 3.

Tab. 4. Data of experimental, predicted and residual for training set (167) of PCB compounds

No	RRT(exp)	RRT(pred)	Residual	NO	RRT(exp)	RRT(pred)	Residual
2	0.1937	0.1906	0.0031	3	0.1975	0.1944	0.0031
4	0.2245	0.2282	−0.0037	5	0.2785	0.2856	−0.0071
7	0.2566	0.2612	−0.0046	8	0.2783	0.2783	0.
9	0.257	0.265	−0.008	10	0.2243	0.2295	−0.0052
12	0.3298	0.3349	−0.0051	13	0.3315	0.3249	0.0066

Tab. 4 – continued:

No	RRT(exp)	RRT(pred)	Residual	NO	RRT(exp)	RRT(pred)	Residual
14	0.2973	0.3212	−0.0239	15	0.3387	0.3287	0.01
17	0.3398	0.3454	−0.0056	18	0.3378	0.3491	−0.0113
19	0.3045	0.3136	−0.0091	20	0.417	0.4199	−0.0029
22	0.4267	0.4199	0.0068	23	0.377	0.3894	−0.0124
24	0.3508	0.3444	0.0064	25	0.3937	0.3955	−0.0018
27	0.3521	0.3638	−0.0117	28	0.4031	0.3955	0.0076
29	0.382	0.3932	−0.0112	30	0.3165	0.3295	−0.013
32	0.3636	0.3638	−0.0002	33	0.4163	0.4188	−0.0025
34	0.3782	0.3917	−0.0135	35	0.4738	0.4692	0.0046
37	0.4858	0.4692	0.0166	38	0.4593	0.4765	−0.0172
39	0.4488	0.4554	−0.0066	40	0.5102	0.514	−0.0038
42	0.487	0.4869	0.0001	43	0.4587	0.4736	−0.0149
44	0.4832	0.4736	0.0096	45	0.4334	0.4285	0.0049
47	0.4639	0.4625	0.0014	48	0.4651	0.4773	−0.0122
49	0.461	0.4492	0.0118	50	0.4007	0.4137	−0.013
52	0.4557	0.453	0.0027	53	0.4187	0.4026	0.0161
54	0.38	0.3487	0.0313	55	0.5562	0.5471	0.0091
57	0.5155	0.5237	−0.0082	58	0.5267	0.5333	−0.0066
59	0.486	0.4786	0.0074	60	0.5676	0.5471	0.0205
62	0.4685	0.4544	0.0141	63	0.529	0.5237	0.0053
64	0.4999	0.4786	0.0213	65	0.4671	0.4582	0.0089
67	0.5214	0.5275	−0.0061	68	0.504	0.5089	−0.0049
69	0.451	0.4638	−0.0128	70	0.5407	0.5446	−0.0039
72	0.4984	0.5127	−0.0143	73	0.4554	0.4772	−0.0218
74	0.5341	0.5275	0.0066	75	0.4643	0.4638	0.0005
77	0.6295	0.6145	0.015	78	0.6024	0.6108	−0.0084
79	0.5894	0.5826	0.0068	80	0.5464	0.5688	−0.0224
82	0.6453	0.6412	0.0041	83	0.6029	0.6179	−0.015
84	0.5744	0.5728	0.0016	85	0.6224	0.6141	0.0083
87	0.6175	0.6007	0.0168	88	0.5486	0.5385	0.0101
89	0.5779	0.5504	0.0275	90	0.5814	0.5907	−0.0093
92	0.5742	0.5774	−0.0032	93	0.5437	0.5423	0.0014
94	0.5331	0.527	0.0061	95	0.5464	0.5323	0.0141
97	0.61	0.6007	0.0093	98	0.5415	0.5233	0.0182
99	0.588	0.5945	−0.0065	100	0.5212	0.5308	−0.0096
102	0.5431	0.5308	0.0123	103	0.5142	0.5175	−0.0033
104	0.4757	0.4487	0.027	105	0.7049	0.6924	0.0125
107	0.6628	0.6691	−0.0063	108	0.6626	0.6605	0.0021
109	0.6016	0.5887	0.0129	110	0.6314	0.624	0.0074
112	0.5986	0.5925	0.0061	113	0.5862	0.592	−0.0058
114	0.6828	0.6619	0.0209	115	0.6171	0.5887	0.0284
117	0.615	0.5925	0.0225	118	0.6693	0.6728	−0.0035
119	0.5968	0.6091	−0.0123	120	0.6256	0.6409	−0.0153
122	0.6871	0.6886	−0.0015	123	0.6658	0.6962	−0.0304
124	0.6584	0.668	−0.0096	125	0.6142	0.6325	−0.0183
127	0.7078	0.7242	−0.0164	128	0.7761	0.7694	0.0067
129	0.7501	0.7561	−0.006	130	0.7284	0.7279	0.0005
132	0.7035	0.6947	0.0088	133	0.6871	0.7046	−0.0175
134	0.6796	0.6866	−0.007	135	0.6563	0.6713	−0.015
137	0.7329	0.729	0.0039	138	0.7403	0.7279	0.0124
139	0.6707	0.6557	0.015	140	0.6707	0.6504	0.0203
142	0.6848	0.6634	0.0214	143	0.6789	0.6653	0.0136
144	0.6563	0.6424	0.0139	145	0.6149	0.5736	0.0413
147	0.6608	0.6595	0.0013	148	0.6243	0.6271	−0.0028

Tab. 4 – continued:

No	RRT(exp)	RRT(pred)	Residual	NO	RRT(exp)	RRT(pred)	Residual
149	0.6672	0.6595	0.0077	150	0.5969	0.5636	0.0333
152	0.6062	0.5774	0.0288	153	0.7036	0.7083	−0.0047
154	0.6349	0.6309	0.004	155	0.5666	0.5488	0.0178
157	0.8184	0.8158	0.0026	158	0.7429	0.734	0.0089
159	0.7655	0.7754	−0.0099	160	0.7396	0.7136	0.026
162	0.7737	0.7925	−0.0188	163	0.7396	0.7378	0.0018
164	0.7399	0.7474	−0.0075	165	0.692	0.7059	−0.0139
167	0.7814	0.7962	−0.0148	168	0.7068	0.7325	−0.0257
169	0.8625	0.8795	−0.017	170	0.874	0.8843	−0.0103
172	0.8278	0.8428	−0.015	173	0.8152	0.8077	0.0075
174	0.7965	0.8095	−0.013	175	0.7611	0.7695	−0.0084
177	0.8031	0.7924	0.0107	178	0.7537	0.7733	−0.0196
179	0.7205	0.7217	−0.0012	180	0.8362	0.8428	−0.0066
182	0.7653	0.7653	0.	183	0.772	0.7695	0.0025
184	0.7016	0.6737	0.0279	185	0.7848	0.7672	0.0176
187	0.7654	0.7733	−0.0079	188	0.692	0.6774	0.0146
189	0.9142	0.9307	−0.0165	190	0.874	0.8589	0.0151
192	0.8269	0.827	−0.0001	193	0.8397	0.8612	−0.0215
194	0.962	0.981	−0.019	195	0.9321	0.9359	−0.0038
197	0.8293	0.829	0.0003	198	0.8845	0.8944	−0.0099
199	0.8494	0.8428	0.0066	200	0.8197	0.8157	0.004

The statistical results of training and external validation of the model are shown in Tab. 6.

Before we begin to investigate stationary phase types and their interaction with analyte molecules, it is essential to understand the concept of molecular polarity and dipole interactions. These interactions form the basis of fundamental adsorption mechanisms that cause analyte retention in gas chromatography (GC). We also classify GC stationary phase types according to their polarity (non-polarity) and so a good understanding is very important.

Tab. 5. Predicted and residual relative retention time for test set (42) of PCB compounds

No	RRT(exp)	RRT(pred)	Residual	NO	RRT(exp)	RRT(pred)	Residual
1	0.1544	0.1441	0.0103	106	0.668	0.6619	0.0061
6	0.2709	0.2783	−0.0074	111	0.6183	0.6371	−0.0188
11	0.3238	0.3249	−0.0011	116	0.6132	0.5793	0.0339
16	0.3625	0.3697	−0.0072	121	0.5518	0.5772	−0.0254
21	0.4135	0.4128	0.0007	126	0.7512	0.7561	−0.0049
26	0.3911	0.3993	−0.0082	131	0.6853	0.6828	0.0025
31	0.4024	0.3993	0.0031	136	0.6257	0.6078	0.0179
36	0.4375	0.4554	−0.0179	141	0.7203	0.7156	0.0047
41	0.499	0.4969	0.0021	146	0.6955	0.7046	−0.0091
46	0.445	0.4232	0.0218	151	0.6499	0.6461	0.0038
51	0.4242	0.3988	0.0254	156	0.8105	0.8073	0.0032
56	0.5676	0.5652	0.0024	161	0.6968	0.7021	−0.0053
61	0.5331	0.5277	0.0054	166	0.7572	0.7136	0.0436
66	0.5447	0.5408	0.0039	171	0.8089	0.811	−0.0021
71	0.4989	0.5091	−0.0102	176	0.7305	0.7179	0.0126
76	0.5408	0.5471	−0.0063	181	0.7968	0.7806	0.0162
81	0.6149	0.6108	0.0041	186	0.7416	0.6985	0.0431
86	0.6105	0.6118	−0.0013	191	0.8447	0.8574	−0.0127
91	0.5549	0.5457	0.0092	196	0.8938	0.9207	−0.0269
96	0.5057	0.4636	0.0421	201	0.8875	0.9073	−0.0198
101	0.5816	0.5812	0.0004	206	1.0103	1.0326	−0.0223

Polysiloxanes are the most common stationary phases. They are available in the greatest variety and are the most stable, robust and versatile. Standard Polysiloxanes are characterized by the repeating siloxane backbone. Each silicon atom contains two functional groups. The type and number of the groups distinguish each stationary phase and its properties. The most basic polysiloxane is 100% methyl substituted. When other groups are present, the number is indicated as the percentage of the total number of groups. For example, SE-54 contains 5% phenyl groups and 95% methyl groups.

In GC, retention of solute molecules occurs due to stronger interaction with the stationary phase than the mobile phase. In GC, the situation is unique in that the chemical interaction with the mobile phase is very small indeed, therefore the interactions between the analyte molecules and the stationary phase are of great importance. In GC, the interaction between the analyte and stationary phase can be divided into three broad categories: dispersive interactions, dipole interactions, and hydrogen bonding.

Dispersive interactions are most difficult to describe and visualize, as they are caused by charge fluctuations that occur throughout a molecule that arise from electron/nuclei vibrations. The fluctuations are random in nature and are basically a statistical effect. Every molecule has a number of arrangements of nuclei and electrons having dipole moments that fluctuate, resulting in an overall molecular charge of zero. However, at any instant in time the dipoles are capable of interacting with other instantaneous dipoles of other molecules. Dispersive forces are ubiquitous and must arise in all molecular interactions. They can themselves occur in isolation, but are always present even when other types of interaction dominate.

Tab. 6. Statistical parameters of QSRR-MLR model

Training set	167
Test set	42
Multiple correlation coefficient (train set)	$R = 0.9972$, $R^2 = 0.9945$
Fischer's criterion (train set)	$F = 4024.77$
Standard deviation (train set)	$SD = 0.0137$
Root mean-squared error (train set)	$\mathrm{RMSE} = 0.0134$
Mean absolute error (train set)	$\mathrm{MAE} = 0.0108$
Squared correlation coefficient of leave-one-out cross-validation	$Q^2_{\mathrm{LOO}} = 0.9938$
Standard deviation error of prediction (test set)	$\mathrm{SDEP} = 0.0139$
Squared correlation coefficient (test set)	$Q^2_{Ext} = 0.9913$
Root mean-squared error (test set)	$\mathrm{RMSE} = 0.0169$
Standard deviation (test set)	$\mathrm{SD} = 0.0173$

There are two distinctive classes of dipole-dipole interaction, those between two species containing a permanent dipole (dipole- dipole interactions) and those between a molecule possessing a permanent dipole and polarizable molecule (dipole-induced dipole interactions). Dipole-dipole interactions can be very strong and occur between molecules with permanent dipole. However, the strength of the dipole-dipole interaction will far exceed any dispersive interactions that occur.

Dipole-induced dipole interactions occur when a molecule containing a permanent dipole approaches a molecule that is polarizable; most commonly these molecules would contain π-electron systems. The strength of this interaction lies between dispersive and dipole-dipole interactions.

This study shows that polar molecules (more Cl atoms) seem to be better retained onto the stationary phase than non-polar molecules (less Cl atoms). Retention onto the stationary phase mainly dependent to Van der Waals forces (dispersive interactions) and dipole-induced dipole interactions (molecular structure- stationary phase of SE-54). Thus in the QSRR here, one sees a general increase in retention times as molecular size and molecular polarity increase, reflected in fragments. Thus in a homologous series such as the PCBs, RRT increases with increasing of Cl atoms and molecular size.

IV. CONCLUSION

In this work, the MLR modeling method was used to study the quantitative structure-retention relationship of RRT on SE-54 stationary phase for a PCBs data set. We can conclude that: firstly, the prediction results indicate that the multi-linear regression modeling method can improve the prediction accuracy significantly for this large data set; secondly, the models developed in this work provide an accurate model that can be used to predict the RRT from the molecular structure only. Physical adsorption onto the stationary phase mainly involves Van der Waals forces and polarity interactions. In this paper, a new QSRR model has been developed for predicting the RRT of PCBs congener from the molecular structure alone. The obtained results show that the MLR method could model the relationship between RRT and their sub-structural fragmental. By performing model validation, it can be concluded that the presented model is a valid model and can be effectively used to predict the RRT of PCBs with an accuracy approximating the accuracy of experimental RRT determination. It can be reasonably concluded that the proposed model would be expected to predict RRT for the test set for which experimental values are unknown. The main advantages of fragment descriptors lie in the simplicity of their computation, easiness of their interpretation as well as efficiency of their applications in similarity searches and SAR/QSAR/QSPR modeling.

Acknowledgement

I wish to thank Prof. V. P. Solovev and Prof A. Varnek for their precious help in use of ISIDA-QSPR software.

References

[1] C.J. Halsall, R.G.M. Lee, P.J. Coleman, et al., *PCBs in UK urban air*, Environ. Sci. Technol., **29**, 2368-2376 (1995).

[2] R.F. Herrick, D.J. Lefkowitz and G.A. Weymouth, *Soil contamination from PCB containing buildings*, Environ. Health Persp., **115**, 173-175 (2007).

[3] S. Bayen, E. Koroleva, H.K. Lee, et al., *Persistent organic pollutants and heavy metals in typical sea foods consumed in Singapore*, J. Toxicol. Environ. Health, Part A **68**, 151-166 (2005).

[4] M.P. Simmonds, K. Haraguchi, T. Endo, et al., *Human health significance of organochlorine and mercury contaminants in Japanese whale meat*, J. Toxicol. Environ. Health, Part A, 65, 1211-1235(2002).

[5] D.G. Wang, M. Yang, H.L. Jia, et al., *Levels, distributions and profiles of polychlorinated biphenyls in surface soils of Dalian*, *China*, Chemosphere, **73**, 38-42 (2008).

[6] S. Ohta, H. Tokusawa, T. Nakao, et al., *Global contamination of coplanar polybrominated/chlorinated biphenyls (Co-PXBs) in the market fishes from Japan*, Chemosphere, **73**, 31-38 (2008).

[7] J. She, A. Holden, T.L. Adelsbach, et al., *Concentrations and time trends of polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) in aquatic bird eggs from San Francisco Bay, CA 2000-2003*, Chemosphere, **73**, 201-209 (2008).

[8] A.S. Souza, J.P.M. Torres, R.O. Meire, et al., *Organochlorine pesticides (OCs) and polychlorinated biphenyls (PCBs) in sediments and crabs (Chasmagnathus granulata, Dana, 1851) from mangroves of Guanabara bay, Rio de Janeiro state, Brazil*, Chemosphere, **73**, 186-192 (2008).

[9] C. Mori, H. Fukata, K. Sakurai, et al., *Strong correlation between the concentration of dioxins and total PCBs in current Japanese people*, Chemosphere, **73**, 235-238 (2008).

[10] Y. Masuda, A. Schecter, and O. Papke, *Concentrations of PCBs, PCDFs and PCDDs in the blood of Yusho patients and their toxic equivalent contribution*, Chemosphere, **37**, 1773-1780 (1998).

[11] S.L. Schantz, *Developmental neurotoxicity of PCBs in humans: what do we know and where do we go from here?,* Neurotoxicol Teratol **18**, 217-227 (1996).

[12] M.D. Erickson, *Analytical chemistry of PCBs*, 2nd edn. CRC Press LLC, Boca Raton, FL, USA 1997.

[13] E.A. Castro, A. A. Toropov, A.I. Nesterova ,et al., *QSPR modeling aqueous solubility of polychlorinated biphenyls by optimization of correlation weights of local and global graph invariants*, Central European Journal of Chemistry, **2**(3), 500-523 (2004).

[14] B. Wei, S. Xie, M. Yu, et al., *QSPR-based prediction of gas/particle partitioning of polychlorinated biphenyls in the atmosphere*, Chemosphere, **66**(10), 1807-1820 (2007).

[15] J. F. Niu, Z. F. Yang, Z. Y. Shen, et al., *QSPRs for the prediction of photodegradation half-life of PCBs in n-hexane*, SAR QSAR Environ. Res., **17**(2), 173-182 (2006).

[16] J. Padmanabhan, R. Parthasarathi, V. Subramanian, et al., *QSPR models for polychlorinated biphenyls: n-Octanol/water partition coefficient*, Bioorg. Med. Chem. Lett., **14**(4), 1021-1028 (2006).

[17] L. Jantschi, S. Bolboacă, *Molecular Descriptors Family on Structure Activity Relationships 6. Octanol-Water Partition Coefficient of Polychlorinated Biphenyls*, Leonardo El. J. Pract. Technol. **8**, 71-86 (2006).

[18] S. Puri, J. S. Chickos and W.J. Welsh, *Three-dimensional quantitative structure – Property relationship (3D-QSPR) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): Enthalpy of vaporization*, J. Chem. Inf. Comp. Sci. **42**(2), 299-304 (2002).

[19] J. Padmanabhan, R. Parthasarathi, V. Subramanian, et al., *Using QSPR models to predict the enthalpy of vaporization of 209 polychlorinated biphenyl congeners*, QSAR Comb. Sci. **26**(2), 227-237 (2007).

[20] S. Puri, J.S. Chickos and W.J. Welsh, *Three-dimensional quantitative structure – Property relationship (3D-QSPR) models for prediction of thermodynamic properties of polychlorinated biphenyls (PCBs): Enthalpy of sublimation*, J. Chem. Inf. Comp. Sci. **42**(1), 109-116 (2002).

[21] J. Devillers and Z. Fresenius, A simple method for the prediction of the GLC retention times of all the 209 PCB congeners. Anal. Chem. **332** (1), 61-62 (1988).

[22] M.N. Hasan and P.C. Jurs, *Computer-assisted prediction of gas chromatographic retention times of polychlorinated biphenyls*, Anal. Chem., **60**(10), 978-982 (1988).

[23] S. Liu, Y. Liu, D. Yin, et al., *Prediction of chromatographic relative retention time of polychlorinated biphenyls from the molecular electronegativity distance vector*, J. Sep. Sci. **29**(2), 296-301(2006).

[24] Y. Ren, H. Liu, X. Yao, et al., *An accurate QSRR model for the prediction of the GC×GCTOFMS retention time of polychlorinated biphenyl (PCB) congeners*, Anal. Bioanal. Chem. **388** (1), 165-172 (2007).

[25] S. Bowadt, H. Skejoandresen, L. Montanarella, et al., *Hrgc separations of 160 chlorobiphenyls in technical mixtures on 4 polar narrow-bore columns*, Int. J. Environ. Anal. Chem. **56**(2), 87-107(1994).

[26] V. Gajduskova and R. Uldrich, *Analysis of specific polychlorinated biphenyl congeners for the examination of raw-materials and foodstuffs of animal origin*, Vet. Med. **37**, 471-478 (1992).

[27] M. Bolgar, J. Cunningham, R. Cooper, et al.,, Physical, spectral and chromatographic properties of all 209 individual PCB congeners, Chemosphere **31**, 2687-2705 (1995).

[28] G. Castello and G. Testini, *Determination of retention indices of polychlorobiphenyls by using other compounds detectable by electron-capture detection or selected polychlorobiphenyls as the reference series*, J. Chromatogr. A **741**, 241-249 (1996).

[29] M. D. Mullin, C.M. Pochini, S. McCrindle, et al., *High-resolution PCB analysis: synthesis and chromatographic properties of all 209 PCB congeners*, Environ. Sci. Technol., **18**,468-476 (1994).

[30] A. Robbat Jr., G. Xyrafas and D. Marshall,*Prediction of gas chromatographic retention characteristic of polychlorinated biphenyls*, Anal. Chem. **60**, 982-985 (1988).

[31] H. A. J. Govers and P. de Voogt, Gas chromatographic derivation of the solubility parameters of polychbrinated biphenyls with the inclusion of cis-trans and optical isomerism and orientational disorder, SAR QSAR Environ. Res. 3, 315-324 (1995).

[32] J. Ghasemi and S. Saaidpour, *QSPR prediction of aqueous solubility of drug-like organic compounds*, Chem. Pharm. Bull. **55**, 669-674 (2007).

[33] J. Ghasemi, S. Saaidpour and S.D. Brown, *QSPR study for estimation of acidity constants of some aromatic acids derivatives using multiple linear regression (MLR) analysis*, J. Mol. Struct. (Theochem.) **805,** 27-32 (2007).

[34] J. Ghasemi and S. Saaidpour, *Quantitative structure-property relationship study of n-octanol- water partition coefficients of some of diverse drugs using multiple linear regression*, Anal. Chim. Acta **604**, 99-106 (2007).

[35] J. Ghasemi and S. Saaidpour, *QSRR prediction of the chromatographic retention behavior of painkiller drugs*, J. Chromatogr. Sci. **47,** 156-163 (2009).

[36] J. Ghasemi and S. Saaidpour, *Artificial neural network-based quantitative structural property relationship for predicting boiling points of refrigerants*, QSAR Comb. Sci., **28,** 1245-1254 (2009).

[37] S. Saaidpour, *Prediction of drug lipophilicity using back propagation artificial neural network modeling*, Orient. J. Chem. **30**(2), 793-802(2014).

[38] S. Saaidpour, A. Bahmani and A. Rostami, *Prediction the normal boiling points of primary, secondary and tertiary liquid amines from their molecular structure descriptors*, CMST **21**(4) 201-210 (2015).

[39] S. Khaledian and S. Saaidpour, *Quantitative structure-property relationship modelling of distribution coefficients (logd7.4) of diverse drug by sub-structural molecular fragments method*, Orient. J. Chem. **31**(4), 1969-1976(2015).

[40] S. Saaidpour, *Quantitative modeling for prediction of critical temperature of refrigerant compounds*, Phys. Chem. Res. **4**(1), 61-71(2016).

[41] S. Saaidpour, S. A. Zarei and F. Nasri, *QSPR study of molar diamagnetic susceptibility of diverse organic compounds using multiple linear regression analysis*, Pak. J. Chem. **2**(1),1-12(2012).

[42] P. Gramatica, N. Navas and R. Todeschini, *3D-Modelling and Prediction by WHIM Descriptors. Part 9. Chromatographic Relative Retention Time and Physico-Chemical Properties of Polychlorinated Biphenyls (PCBs)*, Chemom. Intell. Lab. Syst. **40**, 53-63(1998).

[43] V. P. Solovev, A.Varnek and G. Wipff, *Modeling of Ion Complexation and Extraction Using Substructural Molecular Fragments*, J. Chem. Inf. Comput. Sci. **40**, 847-858 (2000).

[44] A. Varnek, G. Wipff and V.P. Solovev, *Towards an Information System on Solvent Extraction*, Solvent Extr. Ion Exc. **19**,791-837 (2001).

[45] A. Varnek, G. Wipff, V. P. Solovev, et al., *Assessment of the macrocyclic effect for the complexation of crown-ethers with alkali cations using the substructural molecular fragments method*, J. Chem. Inf. Comput. Sci. **42**(4), 812-829 (2002).

[46] V. P. Solovev and A. Varnek, *Anti-HIV activity of hept, tibo and cyclic urea derivatives: structure-property studies, focused combinatorial library generation and hits selection using substructural molecular fragments method*, J. Chem. Inf. Comp. Sci. **43**(5), 1703-1719(2003).

[47] V. P. Solovev and A. Varnek, *Structure-property modeling of metal binders using molecular fragments*, Russ. Chem. Bull. **53,**1434-1445(2004).

[48] A. Varnek and V. P. Solovev, *"In Silico" design of potential anti-hiv actives using fragment descriptors*, Comb. Chem. High T. Scr. **8**(5), 403-416 (2005).

[49] A.Varnek, D. Fourches, F. Hoonakker, et al., *Substructural fragments: an universal language to encode reactions, molecular and supramolecular structures*, J. Comput. Aided Mol. Des. **19**, 693-703 (2005).

[50] A. Dalby, J. G. Nourse, W. D. Hounshell, et al., *Description of several chemical structure file formats used by computer programs developed at molecular design limited*, J. Chem. Inf. Comput. Sci. **32**, 244-255 (1992).

Saadi Saaidpour is an Assistant Professor at the Department of Chemistry, IAU Sanandaj branch, Sanandaj, Iran. In 2004, he received his MSc degree in Analytical Chemistry and in 2008 he received his PhD degree in Analytical Chemistry and Chemometric at Razi University, Kermanshah, Iran. His research interests concern chemometrics methods, computational chemistry and QSAR & QSPR studies.