CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2005**, *4*, 882–910 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of
Molecular Design

December 2005, Volume 4, Number 12, Pages 882–910

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Danail Bonchev on the occasion of the 65th birthday

Highly Correlating Distance–Connectivity Based Topological Indices 3: PCR and PC–ANN Based Prediction of the Octanol–Water Partition Coefficient of Diverse Organic Molecules

Mojtaba Shamsipur,[1] Raoof Ghavami,[2] Bahram Hemmateenejad,[2] and Hashem Sharghi [2]

[1] Department of Chemistry, Razi University, Kermanshah, Iran
[2] Department of Chemistry, Shiraz University, Shiraz, Iran

Received: January 18, 2005; Revised: July 26, 2005; Accepted: September 30, 2005; Published: December 31, 2005

Citation of the article:
M. Shamsipur, R. Ghavami, B. Hemmateenejad, and H. Sharghi, Highly Correlating Distance–Connectivity Based Topological Indices 3: PCR and PC–ANN Based Prediction of the Octanol–Water Partition Coefficient of Diverse Organic Molecules, *Internet Electron. J. Mol. Des.* **2005**, *4*, 882–910, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Highly Correlating Distance–Connectivity Based Topological Indices 3: PCR and PC–ANN Based Prediction of the Octanol–Water Partition Coefficient of Diverse Organic Molecules

Mojtaba Shamsipur,[1,*] Raoof Ghavami,[2] Bahram Hemmateenejad,[2] and Hashem Sharghi [2]

[1] Department of Chemistry, Razi University, Kermanshah, Iran
[2] Department of Chemistry, Shiraz University, Shiraz, Iran

Received: January 18, 2005; Revised: July 26, 2005; Accepted: September 30, 2005; Published: December 31, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (12), 882–910

Abstract

Motivation. Recently, we proposed some new topological indices (Shamsipur indices) based on the distance sum and connectivity of a molecular graph for use in QSAR/QSPR studies. The aim of this study is to examine the ability of the proposed Sh indices in QSPR study of the n–octanol/water partition coefficients (logP) of a diverse set of organic compounds by means of principal component regression (PCR) and principal component–artificial neural network (PC–ANN) modeling methods combining with two factor selection procedures named eigenvalue ranking (EV), and correlation ranking (CR). Experimental values for the partition coefficient ranging from –0.66 (methanol) to 8.16 (2,2',3,3',4,5,5',6,6'–PCB) have been collected from literature for 379 organic compounds with a wide variety of functional groups containing C, H, N, O, and all halogens.

Method. Ten different Sh indices (Sh1 through Sh10) were calculated for each molecule by different combination of the connectivity and distance sum vectors. The Sh topological descriptor data matrix was subjected to principal component analysis for the reduced the dimensionality of a data set and the most significant factors or principal components (PC) were extracted. Both the linear and nonlinear modeling methods were employed for predicting the logP of an extensive set of organic compounds including several structurally diverse groups of compounds (alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aliphatic alcohols, ethers, esters, aldehydes, ketones, carboxylic acids, amines, aromatic hydrocarbons, halogenated hydrocarbons and some polychlorinated biphenyls (PCBs)). Principal component regression and PC–ANN were used as linear and nonlinear modeling methods, respectively.

Results. Principal component analysis of the Sh data matrix showed that the seven PCs could explain 99.97% of variances in the Sh data matrix. The extracted PCs were used as the predictor variables (input) for PCR and ANN (PN–ANN) models. The ANN model could explain 97.98% of variances in the logP data, while the value obtained from PCR procedures were 80.76%. Indeed, linear (MLR) and nonlinear (MLR–ANN) modelings by the use of original Sh indices were performed for comparison. The respective square of correlation coefficients of the prediction obtained by the MLR, PCR, MLR–ANN and PC–ANN are 0.7431, 0.7857, 0.9377 and 0.9626, and the respective standard errors are 0.783, 0.689, 0.361, and 0.281.

Conclusions. Some newly proposed topological indices (Sh indices) has been applied to predict partition

Dedicated on the occasion of the 65th birthday to Danail Bonchev.

* Correspondence author; phone: +98–831–4223307; fax: +98–831–4228439; E–mail: mshamsipur@yahoo.com.

coefficient of a large set of organic compounds. The results of this project showed that factor selection by correlation ranking gives superior results relative to those obtained by eigenvalue ranking. PCR analysis of the data showed that proposed Sh indices could explain about 80% of variations in the logP data; while the variations explained by the ANN modeling were more than 96%. These results confirm the suitability of the indices in QSPR analysis of the lipophilicity data. The Sh indices were calculated in a simple and fast manner and, in comparison with some previously reported QSPR models, produced better results.

Keywords. Topological indices; quantitative structure–property relationships; QSPR; principal component; principal component regression; artificial neural network; correlation ranking; partition coefficient.

Abbreviations and notations

ANN, artificial neural network	PC–ANN, principal component– artificial neural network
FF–ANN, feed–forward– artificial neural network	PLS, partial least squares
PCR, principal component regression	QSAR, quantitative structure–activity relationships
CR, correlation ranking	QSPR, quantitative structure–property relationships
EV, eigenvalue ranking	PCA, principal component analysis
MLR, multilinear regression	

1 INTRODUCTION

Quantitative structure–activity/property relationships (QSAR/QSPR), mathematical equations relating chemical structure to the biological activities or physicochemical properties, have information that is useful for drug design and medicinal chemistry [1]. A major step in constructing the QSAR/QSPR models is to find one or more molecular descriptors that represent variation in the structural property of the molecules by a number. Topological Indices (TIs) are a convenient means of translating chemical constitution into numerical values, which can be used for correlation with physical properties and biological activities. The use of graph invariants in QSAR/QSPR studies has become of major interest in recent years. There are more than 100 kinds of topological indices developed to date [2]. Previously, we proposed a set of new topological indices, named as shamsipur indices (Sh_1 – Sh_{10} indices) [3,4] and used them for prediction of different physical and thermodynamic functions of a large number of alkanes and alkenes isomers.

The octanol–water partition coefficient (K_{ow}=C_o/C_w), defined for dilute solutions as the molar concentration ratio of a single species between *n*–octanol and water phases at equilibrium, is an important physicochemical property, as it provides a useful quantitative parameter for representing the lipophilic/hydrophilic nature of the substance. Its logarithm (logP), after its first usage as a measure of lipophilicity by Hansch and Leo [5], is probably the most informative physicochemical property in medicinal chemistry and is widely used in QSAR/QSPR studies. Because of difficulties in the experimental determination of this parameter and the impossibility of its determination for unsynthesized compounds, there is a strong interest in predicting logP values by theoretical methods in order to be independent from expensive and time–consuming measurements. At the present time, the QSPR/QSAR is a successful strategy for prediction of logP based on modeling between calculated descriptors from molecular structures and logP [6–10].

In general, development of a QSPR involves three steps: structural encoding, feature selection, and model building. Structural encoding involves the use of numerical descriptors to encode the

structural features of a compound. Feature selection is then employed to determine which subset of the descriptors best relates to the property of interest. Models built from the best subset of descriptors from a direct link between descriptors and the property of interest. Finally, validation determines the level of the model's predictive capabilities for unknown compounds. In QSAR/QSPR studies, a regression model of the form $\mathbf{y} = \mathbf{X}\,\mathbf{b} + \mathbf{e}$ may be used to describe a set of predictor variables (**X**) with a predicted variable (**y**) by means of a regression vector (**b**). However, the collinearity, which often existed between independent variables, creates a severe problem in certain types of mathematical treatment such as matrix inversion [11]. A better predictive model can be obtained by ortogonalization of the variables by means of principal component analysis (PCA) and the consequent method is called principal component regression (PCR) [12–14]. In order to reduce the dimensionality of the independent variable space, a limited number of principal components (PCs) are used and therefore a major question will arise after the PCA is how many and which PCs constitute a good subset for predictive purposes? Hence, the selection of significant and informative PCs is the main problem in almost all PCA–based calibration methods. Therefore, different methods including eigenvalue ranking, correlation ranking and genetic algorithm have been addressed to select the significant PCs for calibration purposes [15,16].

Because of the complexity of the relationships existed between the activity/property of the molecules and the structures, nonlinear modeling methods are often used to model the structure–activity/property relationships. Artificial neural networks (ANN) are nonparametric nonlinear modeling techniques that have attracted increasing interest in recent years [17–19]. Nonlinear multivariate maps use a nonlinear transformation of the input variable space to project inputs onto the designated attribute values in an output space. The principal component–artificial neural network (PC–ANN), which combines the PCA with ANN, is another version of the PCR. PC–ANN that models the nonlinear relationships between the PCs and dependent variable was proposed by Gemperline et al. to improve the training speed and decrease the overall calibration error [20]. In this method, the input data are subjected to PCA before being introduced into the neural network and the most significant principal components of the original data matrix are selected and used as ANN input [16–18].

The aim of this study is to examine the ability of the proposed Sh indices in QSPR study of the logP. Both the linear and nonlinear modeling methods were employed for predicting the logP of an extensive set of organic compounds including several structurally diverse groups of compounds (alkanes, alkenes, alkynes, cycloalkanes, cycloalkenes, aliphatic alcohols, ethers, esters, aldehydes, ketones, carboxylic acids, amines, aromatic hydrocarbons, halogenated hydrocarbons and some polychlorinated biphenyls (PCBs)). Principal component regression and PC–ANN were used as linear and nonlinear modeling methods, respectively. Finally, the prediction capabilities of both the linear and nonlinear approaches are tested explicitly by application of the models to subsets of compounds excluded from the calibration set.

2 MATERIALS AND METHODS

2.1 LogP Data

The data set of *n*–octanol/water partition coefficients of diverse organic compounds, was recompiled from several literature sources [21–26]. The choice was based on maximum diversity of the structure of compounds and the numerical values of *n*–octanol/water partition coefficients. The final set of 379 divers organic compounds was representative for all major classes of organic compounds containing C, H, O, N, F, Cl, Br, and I, and included saturated and unsaturated hydrocarbons, halogenated hydrocarbons, polychlorinated biphenyls (PCBs), esters, aldehydes, organic acids, alcohols, ethers, amines, and aromatic compounds (Table 1).

Table 1. The experimental partition coefficient of the data set of 379 organic compounds used in this study and the predicted values by CR–PCR and PC–CR–ANN models

No.	Compound name			CR–PCR		PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
1	Cyclooctane	1	3.28	2.80	–0.48	3.47	0.19
2	2,4,5–PCB	12	5.60	4.95	–0.65	5.41	–0.19
3	Dimethylbenzylamine	10	1.98	3.22	1.24	2.33	0.35
4	1–Heptene	2	3.99	2.42	–1.57	3.51	–0.48
5	Diisopropyl ether	4	1.52	1.64	0.12	1.53	0.01
6	p–Chlorotoluene	12	3.33	2.84	–0.49	3.38	0.05
7	Ethyl iso–propyl ether	4	1.33	1.22	–0.11	1.09	–0.24
8	n–Hexane	1	3.00	2.15	–0.85	3.12	0.12
9	3–Pentanone	7	0.99	1.13	0.14	0.89	–0.10
10	1,2,3,4–Tetrachlorobenzene	12	4.55	3.66	–0.89	4.54	–0.01
11	1–Nonene	2	5.15	3.58	–1.57	4.60	–0.55
12	1–Bromohexane	12	3.80	3.39	–0.41	3.80	0.00
13	Toluene	11	2.73	2.30	–0.43	2.83	0.10
14	Styrene	11	3.05	2.56	–0.49	2.91	–0.14
15	Diphenylamine	10	3.44	3.90	0.46	3.27	–0.17
16	tert–Butylamine	10	0.40	1.39	0.99	0.39	–0.01
17	Iodoethane	12	2.00	1.82	–0.18	2.11	0.11
18	Methyl propionate	8	0.73	0.84	0.11	0.90	0.17
19	2,2,3–Trimethyl–3–pentanol	5	1.99	2.93	0.94	2.33	0.34
20	1,2,3,5–Tetrachlorobenzene	12	4.65	3.71	–0.94	4.57	–0.08
21	2–Methyl–2–hexanol	6	1.87	2.47	0.60	1.81	–0.06
22	1,2,4–Trimethylbenzene	11	3.63	3.12	–0.51	3.61	–0.02
23	3–Hexanol	5	1.61	1.93	0.32	1.62	0.01
24	2,2',4,4',6,6'–PCB	12	7.00	5.94	–1.06	6.73	–0.27
25	Ethanal	6	0.45	0.38	–0.07	–0.17	–0.62
26	1–Bromoheptane	12	4.36	3.99	–0.37	4.33	–0.03
27	Trifluoromethane	12	0.64	–1.20	–1.84	0.86	0.22
28	N–methylaniline	10	1.71	2.46	0.75	1.49	–0.22
29	Ethyl acetate	8	0.73	0.76	0.03	0.89	0.16
30	n–Heptanol	5	2.34	2.57	0.23	2.29	–0.05
31	2–Hexanone	7	1.38	1.52	0.14	1.53	0.15
32	Ethylamine	10	–0.13	–0.15	–0.02	–0.19	–0.06
33	Bromobenzene	12	2.99	3.03	0.04	3.12	0.13
34	Hexanoic acid	9	1.92	1.80	–0.12	1.79	–0.13
35	1,2,3–Trimethylbenzene	11	3.60	3.07	–0.53	3.74	0.14
36	2–Octanone	7	2.37	3.03	0.66	2.72	0.35
37	Cyclohexanol	5	1.23	1.63	0.40	1.18	–0.05

Table 1. (Continued)

No.	Compound name		CR–PCR			PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
38	Methanol	5	–0.66	–0.04	0.62	–0.41	0.25
39	2–Undecanone	7	4.09	4.36	0.27	4.11	0.02
40	Phenyl benzoate	8	3.59	4.24	0.65	3.56	–0.03
41	2–Phenylpropanoic acid	9	1.80	2.97	1.17	1.97	0.17
42	2–Methyl–1–propanol	5	0.61	0.52	–0.09	0.89	0.28
43	Heptanal	6	2.42	2.26	–0.16	2.13	–0.29
44	cis–1,2–Dichloroethene	12	1.86	1.11	–0.75	2.00	0.14
45	n–Propanol	5	0.34	–0.27	–0.61	0.45	0.11
46	2,2–Diphenylacetic acid	9	3.05	4.67	1.62	3.48	0.43
47	Hexadecanoic acid	9	7.17	6.64	–0.53	7.07	–0.10
48	Ethyl decanoate	8	4.73	5.03	0.30	4.96	0.23
49	Pentachlorobenzene	12	5.03	4.97	–0.06	5.10	0.07
50	Ethylbenzene	11	3.15	2.88	–0.27	2.99	–0.16
51	3–Penten–2–ol	5	0.81	1.24	0.43	0.93	0.12
52	1–Methylnaphthalene	11	3.87	3.60	–0.27	4.08	0.21
53	1,2–Dichloropropane	12	2.00	1.90	–0.10	2.27	0.27
54	n–Heptane	1	3.50	2.77	–0.73	3.56	0.06
55	Pentylamine	10	1.49	1.66	0.17	1.49	0.00
56	2–Methyl–2–pentanol	5	1.39	1.92	0.53	1.37	–0.02
57	2–Phenylacetophenone	7	3.18	4.11	0.93	3.11	–0.07
58	3–Methyl–1–butanol	5	1.14	1.08	–0.06	1.43	0.29
59	Propylamine	10	0.48	0.27	–0.21	0.27	–0.21
60	n–Hexanol	5	1.84	1.91	0.07	1.95	0.11
61	2–Fluoroaniline	12	1.26	1.85	0.59	1.72	0.46
62	1–Penten–3–ol	5	0.81	1.29	0.48	0.88	0.07
63	Tribromomethane	12	2.38	3.38	1.00	2.46	0.08
64	Methyl cyclopropyl ether	4	1.24	1.30	0.06	0.98	–0.26
65	Methyl butyrate	8	1.23	1.41	0.18	1.21	–0.02
66	o–Xylene	11	3.12	2.70	–0.42	3.25	0.13
67	Pentachloroethane	12	2.89	2.86	–0.03	2.80	–0.09
68	2–Phenyl–1–propanol	5	1.88	2.76	0.88	2.24	0.36
69	Pentamethylbenzene	11	4.56	3.82	–0.74	4.42	–0.14
70	Butylamine	10	0.75	0.96	0.21	0.78	0.03
71	Acetone	7	–0.24	0.61	0.85	0.01	0.25
72	Ethane	2	1.13	1.39	0.26	1.10	–0.03
73	1,2–Dichloroethane	12	1.48	1.23	–0.25	2.05	0.57
74	Tetrachloroethene	12	2.88	1.96	–0.92	2.82	–0.06
75	Dichloromethane	12	1.25	1.21	–0.04	1.43	0.18
76	m–Toluic acid	9	2.37	2.61	0.24	2.03	–0.34
77	1–Iodopropane	12	2.50	1.95	–0.55	2.45	–0.05
78	Chlorobenzene	12	2.84	2.42	–0.42	2.94	0.10
79	Trimethylamine	10	0.27	1.21	0.94	0.15	–0.12
80	2–Bromobenzoic acid	12	2.20	3.30	1.10	2.34	0.14
81	Benzyl alcohol	5	1.10	2.02	0.92	1.27	0.17
82	Cyclododecanone	7	4.10	4.05	–0.05	3.70	–0.40
83	Octylbenzene	11	6.30	5.13	–1.17	6.30	0.00
84	n–Butylbenzene	11	4.26	3.54	–0.72	4.07	–0.19
85	m–Xylene	11	3.20	2.71	–0.49	3.20	0.00
86	Tertbutylbenzene	11	4.11	3.33	–0.78	3.66	–0.45
87	Ethyl butyrate	8	1.73	2.00	0.27	1.86	0.13
88	1,3,5–Trichlorobenzene	12	4.02	3.36	–0.66	4.13	0.11
89	Tetrafluoromethane	12	1.18	–0.76	–1.94	0.84	–0.34
90	n–Octanol	5	2.84	3.19	0.35	2.69	–0.15
91	1,6–Heptadiyne	3	2.47	1.60	–0.87	1.70	–0.77
92	1,4–Dimethylnaphthalene	11	4.37	3.97	–0.40	4.44	0.07

Table 1. (Continued)

No.	Compound name			CR–PCR		PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
93	2,3,4,5,6–PCB	12	6.30	5.53	–0.77	6.11	–0.19
94	2–Butyne	3	1.46	1.21	–0.25	1.39	–0.07
95	Fluorobenzene	12	2.27	1.51	–0.76	1.64	–0.63
96	Trichloromethane	12	1.97	1.43	–0.54	2.22	0.25
97	1–Bromobutane	12	2.75	2.10	–0.65	2.80	0.05
98	Ethyl propyl ether	4	1.53	1.27	–0.26	1.44	–0.09
99	o–Methylaniline	10	1.32	2.48	1.16	1.61	0.29
100	1,1–Dichloroethane	12	1.79	2.09	0.30	1.93	0.14
101	2–Pentanone	7	0.91	1.01	0.10	0.85	–0.06
102	Methyl decanoate	8	4.41	4.50	0.09	4.29	–0.12
103	2–Pentanol	5	1.14	1.21	0.07	1.14	0.00
104	Cyclopropylamine	10	0.07	1.02	0.95	0.09	0.02
105	Cyclopentene	2	1.75	1.68	–0.07	1.93	0.18
106	Ethylmethylamine	10	0.15	0.41	0.26	0.35	0.20
107	1,1,2,2–Tetrachloroethane	12	2.39	2.71	0.32	2.64	0.25
108	Methyl acetate	8	0.18	0.31	0.13	0.07	–0.11
109	Butyl pentanoate	8	3.23	3.46	0.23	3.34	0.11
110	Dibromomethane	12	2.30	2.47	0.17	1.86	–0.44
111	N,N–dimethylbenzylamine	10	1.98	3.11	1.13	2.28	0.30
112	2–Phenylethanol	5	1.36	2.37	1.01	1.75	0.39
113	1,2,4,5–Tetrachlorobenzene	12	4.51	3.72	–0.79	4.59	0.08
114	2–Hexanol	5	1.61	1.89	0.28	1.69	0.08
115	Ethyl isobutyrate	8	1.53	2.03	0.50	1.92	0.39
116	Methanal	6	0.35	0.89	0.54	–0.20	–0.55
117	3,3–Dimethyl–2–butanol	5	1.19	1.82	0.63	1.28	0.09
118	1,4–Dibromobenzene	12	3.79	3.63	–0.16	3.87	0.08
119	1–Octene	2	4.57	3.02	–1.55	4.03	–0.54
120	n–Propyl–n–butylamine	10	2.12	2.97	0.85	2.44	0.32
121	Iodomethane	12	1.50	2.45	0.95	1.45	–0.05
122	Pentylbenzene	11	4.90	3.91	–0.99	4.67	–0.23
123	2–Methyl–3–pentanone	7	1.09	1.55	0.46	1.13	0.04
124	1,1,1–Trichloroethane	12	2.49	2.47	–0.02	2.30	–0.19
125	Benzyl acetate	8	1.96	3.02	1.06	2.20	0.24
126	1–Pentene	2	2.20	1.26	–0.94	2.27	0.07
127	1–Chlorobutane	12	2.64	1.66	–0.98	2.55	–0.09
128	3–Methyl–2–butanol	5	0.91	1.18	0.27	0.90	–0.01
129	cis–2–Butene	2	2.33	0.90	–1.43	1.82	–0.51
130	3–Phenylpropionic acid	9	1.84	2.78	0.94	2.02	0.18
131	p–Xylene	11	3.25	2.70	–0.55	3.07	–0.18
132	4–Methyl–2–pentanol	5	1.41	1.77	0.36	1.60	0.19
133	n–Propyl isopropyl ether	4	1.83	1.79	–0.04	1.68	–0.15
134	Hexanal	6	1.78	1.61	–0.17	1.70	–0.08
135	Di–tert–butyl ketone	7	3.00	2.95	–0.05	2.93	–0.07
136	Ethyl methacrylate	8	1.94	2.14	0.20	2.02	0.08
137	Acetophenone	7	1.63	2.51	0.88	1.86	0.23
138	1–Fluorobutane	12	2.58	0.00	–2.58	1.86	–0.72
139	sec–Butyl acetate	8	1.53	1.92	0.39	1.62	0.09
140	1–Heptyne	3	2.61	2.14	–0.47	2.60	–0.01
141	Tetrachloromethane	12	2.64	2.63	–0.01	2.55	–0.09
142	1–Iodoheptane	12	4.70	4.81	0.11	4.64	–0.06
143	Methyl acrylate	8	0.80	0.71	–0.09	0.66	–0.14
144	Acetic acid	9	–0.17	0.23	0.40	–0.25	–0.08
145	1,8–Nonadiyne	3	3.06	2.66	–0.40	3.05	–0.01
146	Tetradecanoic acid	9	6.10	5.87	–0.23	6.02	–0.08

Table 1. (Continued)

No.	Compound name		CR–PCR			PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
147	Dimethyl ether	4	0.10	0.04	–0.06	0.15	0.05
148	p–Trifluoromethylaniline	12	2.39	2.33	–0.06	2.48	0.09
149	Methyl iso–propyl ether	4	0.73	0.62	–0.11	0.66	–0.07
150	Fluoromethane	12	0.51	0.04	–0.47	0.39	–0.12
151	Methyl methacrylate	8	1.38	1.35	–0.03	1.49	0.11
152	3–Methyl–3–pentanol	5	1.39	1.97	0.58	1.25	–0.14
153	2,3,4,5–PCB	12	5.72	5.25	–0.47	5.76	0.04
154	Methyl 2–phenylacetate	8	1.83	2.98	1.15	2.18	0.35
155	3–Methyl–2–pentanone	7	1.09	1.70	0.61	1.11	0.02
156	Ethyl formate	8	0.23	0.37	0.14	0.23	0.00
157	3–Methylpentane	1	2.80	2.16	–0.64	2.74	–0.06
158	4–Bromobenzoic acid	12	2.86	3.58	0.72	2.41	–0.45
159	Pentanal	6	1.31	0.95	–0.36	1.18	–0.13
160	2–Methyl–4–penten–3–ol	5	1.11	1.77	0.66	1.19	0.08
161	1,1,2–Trichloroethane	12	2.38	1.97	–0.41	2.30	–0.08
162	iso–Butylamine	10	0.73	1.02	0.29	0.55	–0.18
163	Dibutylamine	10	2.68	3.55	0.87	2.96	0.28
164	Triphenylamine	10	5.74	5.57	–0.17	6.12	0.38
165	N–ethylpiperidine	10	1.88	2.65	0.77	2.22	0.34
166	Dimethylamine	10	–0.38	0.26	0.64	–0.25	0.13
167	Methyl benzoate	8	2.20	2.67	0.47	2.22	0.02
168	1,3–Dibromobenzene	12	3.75	3.58	–0.17	3.77	0.02
169	cis–1,3–Dichloropropene	12	2.03	1.48	–0.55	2.39	0.36
170	Butyric acid	9	0.79	0.76	–0.03	0.58	–0.21
171	2–Aminonaphthalene	10	2.28	3.39	1.11	2.33	0.05
172	n–Propylbenzene	11	3.69	3.20	–0.49	3.52	–0.17
173	m–Methylaniline	10	1.40	2.46	1.06	1.55	0.15
174	1–Pentyne	3	1.83	1.19	–0.64	2.04	0.21
175	2,2,2–Trifluoroethanol	12	0.37	–0.23	–0.60	1.11	0.74
176	5–Hexyne–2–one	7	0.58	1.29	0.71	1.09	0.51
177	Methyl 3–phenylpropionate	8	2.32	3.35	1.03	2.64	0.32
178	2–Propanol	5	0.14	0.09	–0.05	0.18	0.04
179	1,1,2–Trichlorotrifluoroethane	12	3.16	1.34	–1.82	3.20	0.04
180	1–Iodobutane	12	3.00	2.66	–0.34	3.05	0.05
181	2,2',3,3',4,4'–PCB	12	7.00	5.93	–1.07	6.75	–0.25
182	Bromocyclohexane	12	3.20	2.96	–0.24	3.07	–0.13
183	Ethyne	3	0.37	2.09	1.72	0.73	0.36
184	2–Cyclohexen–1–one	7	0.61	1.55	0.94	0.85	0.24
185	Azobenzene	11	3.82	3.98	0.16	3.61	–0.21
186	1,2–Dimethylcyclohexane	1	3.06	2.89	–0.17	3.16	0.10
187	Isopropyl benzoate	8	3.18	3.62	0.44	2.97	–0.21
188	p–Methylaniline	10	1.39	2.45	1.06	1.51	0.12
189	2,4–Dimethyl–3–pentanol	5	1.71	2.36	0.65	1.73	0.02
190	1,2,3,5–Tetramethylbenzene	11	4.10	3.43	–0.67	3.64	–0.46
191	Hexachlorobenzene	12	5.47	4.36	–1.11	5.13	–0.34
192	2,4–Dimethyl–2–pentanol	5	1.67	2.31	0.64	1.74	0.07
193	2–Methyl–2–propanol	5	0.37	0.84	0.47	0.44	0.07
194	*o*–Toluic acid	9	2.32	2.63	0.31	2.09	–0.23
195	3–Pentanol	5	1.14	1.27	0.13	1.04	–0.10
196	Hexylamine	10	2.06	2.33	0.27	1.91	–0.15
197	1,3,5–Trimethylbenzene	11	3.42	3.12	–0.30	3.54	0.12
198	3–Hexanone	7	1.38	1.59	0.21	1.49	0.11
199	Phenanthrene	11	4.52	4.40	–0.12	4.42	–0.10
200	2,3–Ddimethyl–2–butanol	5	1.17	1.78	0.61	1.16	–0.01
201	Ethyl heptanoate	8	3.23	3.56	0.33	3.20	–0.03

Table 1. (Continued)

No.	Compound name		CR–PCR			PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
202	1,2–Dimethylnaphthalene	11	4.31	3.95	–0.36	4.36	0.05
203	2,2',3,3',5,5',6,6'–PCB	12	7.10	6.38	–0.72	7.19	0.09
204	1,3–Cyclohexadiene	2	2.47	1.80	–0.67	2.16	–0.31
205	*n*–Propyl–*sec*–butylamine	10	1.91	2.92	1.01	2.03	0.12
206	1–Fluoropentane	12	2.33	0.77	–1.56	2.46	0.13
207	2,2–Dimethyl–3–pentanol	5	1.69	2.53	0.84	1.73	0.04
208	2–Methylnaphthalene	11	4.00	3.61	–0.39	4.04	0.04
209	Methylcyclohexane	1	2.76	2.43	–0.33	2.91	0.15
210	Benzylamine	10	1.09	2.15	1.06	1.24	0.15
211	Ethyl Nonanoate	8	4.23	4.56	0.33	4.36	0.13
212	2–Heptanone	7	1.79	2.08	0.29	1.97	0.18
213	Cyclopentane	1	2.05	1.63	–0.42	2.04	–0.01
214	*trans*–1,2–Dichloroethene	12	1.93	1.11	–0.82	2.00	0.07
215	Ethyl propionate	8	1.21	1.52	0.31	1.34	0.13
216	Anthracene	11	4.50	4.45	–0.05	4.42	–0.08
217	Ethyl benzoate	8	2.64	3.49	0.85	2.69	0.05
218	*p*–Ethyltoluene	11	3.63	3.34	–0.29	3.61	–0.02
219	Diethylamine	10	0.57	0.99	0.42	0.89	0.32
220	Chloroethylene	12	1.38	1.20	–0.18	1.41	0.03
221	sec–Butylamine	10	0.74	1.12	0.38	0.65	–0.09
222	1–Chloroheptane	12	4.15	3.48	–0.67	4.13	–0.02
223	*p*–Cymene	11	4.10	3.51	–0.59	3.76	–0.34
224	Isopropylbenzene	11	3.66	3.04	–0.62	3.38	–0.28
225	2–Nonanone	7	2.79	3.29	0.50	2.96	0.17
226	2,2–Dimethyl–1–propanol	5	1.36	1.26	–0.10	1.21	–0.15
227	Diiodomethane	12	2.50	3.68	1.18	2.66	0.16
228	Methyl tert–butyl ether	4	1.06	1.28	0.22	1.16	0.10
229	Formic acid	9	–0.54	–0.54	0.00	–0.34	0.20
230	2,3–Dimethyl–2–pentanol	5	1.67	2.44	0.77	1.68	0.01
231	Ethyl acrylate	8	1.32	1.48	0.16	1.17	–0.15
232	Hexamethylbenzene	11	4.69	4.15	–0.54	4.75	0.06
233	2–Aminooctane	10	2.82	3.52	0.70	2.69	–0.13
234	5–Methyl–2–octanone	7	2.92	3.27	0.35	2.58	–0.34
235	*n*–Pentane	1	2.50	1.52	–0.98	2.46	–0.04
236	Hexachloroethane	12	4.00	3.20	–0.80	3.74	–0.26
237	3–Methyl–2–butanone	7	0.56	1.09	0.53	0.71	0.15
238	Iodobenzene	12	3.28	4.14	0.86	3.15	–0.13
239	1–Aminonaphthalene	10	2.25	3.44	1.19	2.60	0.35
240	Methyl butyl ether	4	1.53	1.35	–0.18	1.54	0.01
241	*o*–Ethyltoluene	11	3.53	3.28	–0.25	3.60	0.07
242	2,2',3,3',4,4',6–PCB	12	6.70	6.19	–0.51	7.06	0.36
243	Iso–propylamine	10	0.26	0.64	0.38	0.14	–0.12
244	Methyl sec–butyl ether	4	1.33	1.31	–0.02	1.02	–0.31
245	2–Methylpentane	1	2.80	2.14	–0.66	2.83	0.03
246	trans–Stilbene	11	4.81	4.28	–0.53	4.67	–0.14
247	Cyclohexanone	7	0.81	1.40	0.59	0.93	0.12
248	1,3,5–Cycloheptatriene	2	2.63	2.11	–0.52	2.37	–0.26
249	3–Chloroaniline	12	1.88	2.55	0.67	1.63	–0.25
250	1,2–Dibromobenzene	12	3.64	3.72	0.08	3.58	–0.06
251	Ethyl–iso–propylamine	10	0.93	1.64	0.71	1.24	0.31
252	Chloromethane	12	0.91	0.82	–0.09	1.00	0.09
253	Trichloroethylene	12	2.35	1.68	–0.67	2.26	–0.09
254	2,2',3,3',4,5,5',6,6'–PCB	12	8.16	6.63	–1.53	7.96	–0.20
255	1,2,3,4–tetramethylbenzene	11	4.00	3.46	–0.54	4.16	0.16

Table 1. (Continued)

No.	Compound name		CR–PCR			PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
256	n–Octane	1	4.00	3.36	–0.64	4.06	0.06
257	1–Octyne	3	2.91	2.72	–0.19	3.17	0.26
258	Benzyl benzoate	8	3.97	4.58	0.61	4.24	0.27
259	Bromochloromethane	12	1.41	1.79	0.38	1.72	0.31
260	2–Butanone	7	0.29	0.71	0.42	0.49	0.20
261	2–Methylbutane	1	2.30	1.65	–0.65	2.16	–0.14
262	1–Hexene	2	3.40	1.83	–1.57	3.02	–0.38
263	2,6–PCB	12	5.00	4.54	–0.46	4.85	–0.15
264	o–Chlorotoluene	12	3.42	2.80	–0.62	3.35	–0.07
265	Bromomethane	12	1.19	1.56	0.37	1.12	–0.07
266	Dibutyl ether	4	3.21	3.19	–0.02	2.68	–0.53
267	Dodecanoic acid	9	4.60	4.99	0.39	4.99	0.39
268	*m*–Chlorotoluene	12	3.28	2.82	–0.46	3.36	0.08
269	2–PCB	12	4.52	4.15	–0.37	4.46	–0.06
270	5–Nonanone	7	2.79	3.19	0.40	2.88	0.09
271	Dimethyl–n–butylamine	10	1.70	2.51	0.81	1.78	0.08
272	2,5–PCB	12	5.10	4.58	–0.52	4.94	–0.16
273	Tripropylamine	10	2.79	3.79	1.00	2.94	0.15
274	Pentyl acetate	8	2.23	2.39	0.16	2.36	0.13
275	Benzene	11	2.13	1.93	–0.20	2.45	0.32
276	Vinyl ecetate	8	0.73	0.61	–0.12	0.47	–0.26
277	3–Heptanone	7	1.79	2.39	0.60	1.91	0.12
278	Aniline	10	0.91	1.98	1.07	1.14	0.23
279	1,2,4–Trichlorobenzene	12	3.98	3.30	–0.68	4.11	0.13
280	Benzoic acid	9	1.88	2.12	0.24	1.30	–0.58
281	4–Phenylbutyric acid	9	2.42	3.20	0.78	2.64	0.22
282	Phenyl formate	8	1.26	2.04	0.78	1.24	–0.02
283	2–Methyl–2–butanol	5	0.89	1.41	0.52	0.83	–0.06
284	1–Hexene–3–ol	5	1.31	1.87	0.56	1.54	0.23
285	Benzophenone	7	3.18	4.39	1.21	3.79	0.61
286	2,2',3,3',6,6'–PCB	12	6.70	5.83	–0.87	6.56	–0.14
287	Azulene	11	3.22	3.26	0.04	3.76	0.54
288	Isobutene	2	2.35	1.46	–0.89	1.61	–0.74
289	Propyl formate	8	0.83	0.77	–0.06	0.74	–0.09
290	Phenyl acetate	8	1.49	2.59	1.10	1.86	0.37
291	2–Bromopropane	12	1.90	2.07	0.17	1.98	0.08
292	2,4,6–PCB	12	5.47	4.94	–0.53	5.37	–0.10
293	1–Hexyne	3	2.26	1.60	–0.66	2.14	–0.12
294	Propyl acetate	8	1.24	1.36	0.12	1.33	0.09
295	Hexylbenzene	11	5.52	4.44	–1.08	5.52	0.00
296	2,4–Dimethylpentane	1	3.10	2.60	–0.50	3.05	–0.05
297	3–Bromopropene	12	1.79	1.74	–0.05	2.23	0.44
298	Diphenylcarbinol	5	2.67	4.23	1.56	3.32	0.65
299	Difluoromethane	12	0.20	–1.43	–1.63	0.20	0.00
300	Dichlorodifluoromethane	12	2.16	0.13	–2.03	1.42	–0.74
301	Naphthalene	11	3.35	3.23	–0.12	3.72	0.37
302	p–Toluic acid	9	2.34	2.63	0.29	1.99	–0.35
303	2–Chloroaniline	12	1.90	2.53	0.63	1.67	–0.23
304	n–Butanol	5	0.88	0.46	–0.42	0.95	0.07
305	Butyl acetate	8	1.82	1.90	0.08	1.78	–0.04
306	Methacrylic acid	9	0.93	0.79	–0.14	0.69	–0.24
307	Propionic acid	9	0.33	0.81	0.48	0.09	–0.24
308	trans–Cinnamic acid	9	2.13	2.82	0.69	1.91	–0.22
309	5–Hexene–2–one	7	1.02	1.29	0.27	1.39	0.37
310	5–Methyl–2–hexanone	7	1.88	2.00	0.12	1.74	–0.14

Table 1. (Continued)

No.	Compound name			CR–PCR		PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
311	1–Bromooctane	12	4.89	4.59	–0.30	4.91	0.02
312	1,3–Dimethylnaphthalene	11	4.42	3.95	–0.47	4.40	–0.02
313	Octanoic acid	9	3.05	2.97	–0.08	2.85	–0.20
314	1–Bromopropane	12	2.10	1.53	–0.57	2.24	0.14
315	n–Pentanol	5	1.40	1.21	–0.19	1.62	0.22
316	1,2,4,5–Tetramethylbenzene	11	4.10	3.47	–0.63	3.88	–0.22
317	Butyl methacrylate	8	2.88	2.95	0.07	2.60	–0.28
318	1,5–Hexadiene	2	2.80	1.49	–1.31	2.65	–0.15
319	2–Chloropropane	12	1.90	1.59	–0.31	1.72	–0.18
320	1,1–Dichloroethylene	12	2.13	1.73	–0.40	1.79	–0.34
321	1–Bromopentane	12	3.37	2.77	–0.60	3.34	–0.03
322	Ethyl octanoate	8	3.73	4.09	0.36	3.77	0.04
323	2,4–Dimethyl–3–pentanone	7	1.39	2.04	0.65	1.77	0.38
324	4–PCB	12	4.61	4.26	–0.35	4.62	0.01
325	Isobutylbenzene	11	4.01	3.36	–0.65	3.75	–0.26
326	Cyclohexane	1	2.46	1.90	–0.56	2.46	0.00
327	Methyl iso–butyl ether	4	1.33	1.33	0.00	1.32	–0.01
328	Methyl n–propyl ether	4	1.03	0.66	–0.37	0.85	–0.18
329	2–Methyl–3–pentanol	5	1.41	1.90	0.49	1.41	0.00
330	1,3–Dichlorobenzene	12	3.48	2.89	–0.59	3.46	–0.02
331	Chloroethane	12	1.43	0.96	–0.47	1.37	–0.06
332	Methylamine	10	–0.58	0.00	0.58	–0.44	0.14
333	Salicylic acid	9	2.20	2.37	0.17	1.82	–0.38
334	3,3–Dimethyl–1–butanol	5	1.86	1.68	–0.18	1.75	–0.11
335	1,2,3–Trichloropropane	12	2.63	2.40	–0.23	2.89	0.26
336	Bromoethane	12	1.60	1.48	–0.12	1.70	0.10
337	1,2–Dichlorobenzene	12	3.38	2.88	–0.50	3.44	0.06
338	1,4–Dichlorobenzene	12	3.38	2.94	–0.44	3.49	0.11
339	2,2',4',5–PCB	12	5.73	5.33	–0.40	5.95	0.22
340	Heptylamine	10	2.57	2.96	0.39	2.36	–0.21
341	2–Methyl–1–butanol	5	1.14	1.29	0.15	1.14	0.00
342	1–Chloropropane	12	2.04	1.11	–0.93	1.97	–0.07
343	3–PCB	12	4.58	4.21	–0.37	4.53	–0.05
344	2,2',4,5,5'–PCB	12	6.40	5.63	–0.77	6.38	–0.02
345	3–Ethyl–3–pentanol	5	1.87	2.47	0.60	1.72	–0.15
346	Triethylamine	10	1.45	2.23	0.78	1.65	0.20
347	4–Penten–1–ol	5	1.04	0.90	–0.14	1.17	0.13
348	Isopropyl acetate	8	1.03	1.25	0.22	0.99	–0.04
349	2–Phenylacetic acid	9	1.41	2.40	0.99	1.58	0.17
350	Piperidine	10	0.85	1.51	0.66	0.92	0.07
351	Dipropyl ether	4	2.03	1.93	–0.10	1.82	–0.21
352	Trichlorofluoromethane	12	2.53	0.44	–2.09	1.72	–0.81
353	3–Methyl–3–hexanol	5	1.87	2.51	0.64	1.75	–0.12
354	4–Chloroaniline	12	1.83	2.55	0.72	1.59	–0.24
355	3–Methyl–2–pentanol	5	1.41	1.90	0.49	1.43	0.02
356	Pentanoic Acid	9	1.39	1.23	–0.16	1.28	–0.11
357	1–Aminooctane	10	3.09	3.55	0.46	2.87	–0.22
358	Dipropylamine	10	1.67	2.33	0.66	1.99	0.32
359	Propanal	6	0.59	0.05	–0.54	0.21	–0.38
360	4–Methyl–2–pentanone	7	1.09	1.47	0.38	1.18	0.09
361	Diethyl ether	4	1.03	0.58	–0.45	0.83	–0.20
362	Ethanol	5	–0.32	–0.54	–0.22	–0.16	0.16
363	2,3–Dimethyl–3–pentanol	5	1.67	2.42	0.75	1.69	0.02
364	Decanoic acid	9	4.60	4.03	–0.57	3.93	–0.67

Table 1. (Continued)

No.	Compound name		CR–PCR			PC–CR–ANN	
		subset	exp logP	pred logP	residual	pred logP	residual [a]
365	Cyclohexylamine	10	1.49	2.03	0.54	1.38	–0.11
366	1–Ethylnaphthalene	11	4.40	4.18	–0.22	4.42	0.02
367	Methylcyclopentane	1	2.35	2.10	–0.25	2.31	–0.04
368	trans–2–Butene	2	2.31	0.90	–1.41	1.82	–0.49
369	2–Hexene–4–ol	5	1.31	1.92	0.61	1.47	0.16
370	Cycloheptane	1	2.87	2.37	–0.50	3.05	0.18
371	1–Nonyne	3	3.18	3.29	0.11	3.86	0.68
372	Butanal	6	0.88	0.36	–0.52	0.59	–0.29
373	Methyl–n–butylamine	10	1.33	1.72	0.39	1.48	0.15
374	2–Decanone	7	3.77	3.84	0.07	3.53	–0.24
375	1,2,3–Trichlorobenzene	12	4.04	3.27	–0.77	4.02	–0.02
376	n–Nonanol	5	3.15	3.76	0.61	3.15	0.00
377	3–Bromobenzoic acid	12	2.87	3.42	0.55	2.37	–0.50
378	Methyl 4–phenylbutyrate	8	2.77	3.76	0.99	3.26	0.49
379	Ethyl hexanoate	8	2.73	3.04	0.31	2.68	–0.05

[a] Subsets of compounds: 1, alkanes; 2, alkenes; 3, alkynes; 4, ethers; 5, alcohols; 6, aldehydes; 7, ketones; 8, esters; 9, acids; 10, amines and nitrogen compounds; 11, aromatic compounds; 12, halogenated hydrocarbons and polychlorinated biphenyls. CR–PCR: The compounds used in the calibration set: 1–275 + 333–379. The compounds used in the validation set: 276–332. PC–CR–ANN: The compounds used in the calibration set: 1–275. The compounds used in the validation set: 276–332. The compounds used in the test set: 333–379. residual = pred logP – exp logP

2.2 Sh Topological Indices

Ten different Sh topological indices (Sh1–Sh10) were calculated for each molecule based on the different combinations of the distance sum and connectivity vectors [3,4]. Let $G = \{V,E\}$ be a hydrogen depleted graph of a molecule with n atom, where V is the vertex set and E the edge set [26,27]. For this graph, two vectors including distance sum(s) and connectivity (δ^v) were calculated [26,27]. The new Sh topological indices were calculated by combination of connectivity and distance sum as:

$$Sh_1 = \log(\sum \frac{S_i S_j}{\delta_i \delta_j}) \quad (1)$$

$$Sh_2 = \log(\sum \frac{\delta_i \delta_j}{S_i S_j}) \quad (2)$$

$$Sh_3 = \log(\sum (S_i S_j \delta_i \delta_j)^{-0.5}) \quad (3)$$

$$Sh_4 = \log(\sum (\frac{\delta_i \delta_j}{S_i S_j})^{-0.5}) \quad (4)$$

$$Sh_5 = \sum (S_i S_j + \delta_i \delta_j)^{-0.5} \quad (5)$$

$$Sh_6 = \log(\sum((S_i S_j) + (\delta_i \delta_j)) \quad (6)$$

$$Sh_7 = \sum(\delta_i \delta_j + \log(S_i S_j)) \quad (7)$$

In above equations, the sums were given over all edges (bonds) and i and j represent the two adjacent atoms. Three other proposed indices were calculated based on individual methods: The distance sums and connectivity were collected in two separate column vectors (**S** and **δ**, respectively). The logarithm of the inner product of **S** and **δ** gives the Sh_8 index:

$$Sh_8 = \log(S^T \delta) \quad (8)$$

Post multiplication of **S** by δ^T gives a square matrix (**Sd**). The sum over all entries of **Sd** is Sh_9:

$$Sd = S\delta^T \quad (9)$$

$$Sh_9 = \log(\sum_i \sum_j Sd_{ij}) \quad (10)$$

The Sh_{10} index was calculated by the summation over eigenvalues of the **Sd** matrix. The eigenvalues of **Sd** are calculated by singular–value decomposition (SVD) [28, 29]. In SVD an individual matrix is decomposed to three matrices: (*i.e.*, U, Σ and V)

$$Sd = U\Sigma V^T \quad (11)$$

Σ is a diagonal matrix whose elements are the eigenvalues of **Sd**. **U** and **V** are row and column designate of **Sd**. In our work, **Sd** has only one eigenvalue and other eigenvalues are zeroed. We represented the eigenvalue of **Sd** by ES, so that Sh_{10} is defined as:

$$Sh_{10} = \log(ES) \quad (12)$$

A home–made program (written in MATLAB environment) calculated the Sh indices. The calculated indices were collected in a data matrix with 379×10 dimension. Each chemical is now a point in the 10–dimensional space, $\mathbf{X}^{10}$ (see Table 2).

2.3 Linear Modeling: Principal Component Regression

Due to the some co–linearity between the Sh topological indices, orthogonal transformation of the Sh indices by principal component analysis was performed. The score and loading matrices were calculated by singular value decomposition (SVD) procedure [28,29]:

$$\mathbf{D} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T \quad (13)$$

where **U** and **V** are the orthonormal matrices spanned the respective row and column spaces of the data matrix (**D**). **S** is a diagonal matrix whose elements are the squared root of the eigenvalues. The superscript "T" denotes the transpose of the matrix. The eigenvectors included in **U** are named as

principal components (PC). The PCs of the validation ($\mathbf{D}_v$) and Prediction ($\mathbf{D}_p$) sets were calculated by the equation:

$$\mathbf{U}_{p/v} = \mathbf{D}_{p/v}\, \mathbf{S}^{-1}\, \mathbf{V} \qquad (14)$$

Application of the PCA on the Sh indices data matrix resulted in 10 principal components or factors (F_1–F_{10}). A linear regression model was build between the logP and resulted factors. The best set of factors was selected by the eigenvalue ranking (EV) and correlation ranking (CR) procedures [16–18]. In the EV–PCR procedure, the PCs were entered to the PCR model consecutively based on their decreasing eigenvalue. Once each new factor was entered to the model, the model performances were evaluated by the leave–one–out cross–validation (LOO–CV). In the CR–PCR, the correlation between each one of the extracted PC's with the logP data was determined first. The stepwise entrance of the PCs to the PCR model was based on their decreasing correlation with the logP. Some statistical parameters such as the squared of the correlation coefficient (R^2), squared of the leave–one–out cross validation correlation coefficient (R^2_{CV}), the standard error of estimation (SE), the root–mean–square error (*RMS*) and the Fisher's criterion at the 95% probability level were calculated to estimate the quality of the resulted models.

2.4 Nonlinear Modeling: PC–ANN

To model the logP–Sh indices more accurate, artificial neural network was employed to process the nonlinear relationships between the selected PCs in the previous section and logP data. The PC–ANN model was the same as we reported previously [17,18,30–32]. The totals of 379 compounds were randomly divided to 275 calibration (or training) samples, 57 prediction samples and 47 validation samples. The PCs of the calibration samples were calculated by equation 13 and those of prediction and validation samples were calculated by equation 14. It should be noted that the PCs calculated by SVD are orthonormal vectors (*i.e.*, orthogonal vector that normalized to length one) and, therefore, normalization of the input vectors is not required. The prediction set is a subset of compounds used to help find an optimal set of weights and biases during ANN calibrating, and it is also used to avoid overtraining of the ANN. The ANNs used in this study were fully connected, three layer, feed–forward ANN. The number of neurons in the input layer is equal to the number of PCs selected for the model. The PC's used here were those selected by the CR–PCR and EV–PCR models. The transformed values are then passed to the hidden layer. The input value of a hidden layer neuron is the summation of the products of the weights (neuron connections) times the corresponding outputs of the previous input layer plus a bias term. The ANN model confined to a single hidden layer, because the network with more than one hidden layer would be harder to train. The summation is put through a nonlinear transfer function; here a sigmoid, and then the resulting values are passed to the output layer, which contains a single neuron, which represents the predicted logP values.

Table 2. List of 379 chemical compounds and their calculated Sh indices matrix

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
1	Cyclooctane	2.709	–0.903	–0.602	1.806	0.496	3.318	51.266	2.408	3.311	4.817
2	2,4,5–PCB	3.793	–0.871	–0.782	2.532	0.501	4.662	306.750	3.286	4.484	6.699
3	Dimethylbenzylamine	3.017	–0.654	–0.717	2.046	0.549	3.888	149.750	2.813	3.833	5.737
4	1–Heptene	2.648	–0.851	–0.646	1.729	0.458	3.248	48.601	2.340	3.195	4.750
5	Diisopropyl ether	2.393	–0.403	–0.660	1.551	0.476	2.988	61.065	2.279	3.186	4.899
6	*p*–Chlorotoluene	2.525	–0.276	–0.583	1.754	0.744	3.386	131.285	2.497	3.430	5.138
7	Ethyl iso–propyl ether	2.159	–0.294	–0.621	1.383	0.500	2.711	47.820	2.114	2.952	4.569
8	n–Hexane	2.371	–0.811	–0.575	1.513	0.460	2.796	26.335	2.041	2.845	4.187
9	3–Pentanone	2.144	–0.040	–0.575	1.389	0.620	2.765	79.508	2.193	2.997	4.619
10	1,2,3,4–Tetrachlorobenzene	2.975	–0.422	–0.615	1.999	0.652	3.722	166.706	2.686	3.732	5.613
11	1–Nonene	3.210	–1.203	–0.757	2.065	0.354	3.807	65.393	2.675	3.636	5.406
12	1–Bromohexane	3.005	–1.012	–0.578	1.817	0.404	3.152	32.583	2.221	3.101	4.597
13	Toluene	2.207	–0.136	–0.537	1.593	0.838	3.158	115.241	2.366	3.225	4.803
14	Styrene	2.539	–0.280	–0.590	1.796	0.770	3.484	142.481	2.573	3.487	5.194
15	Diphenylamine	3.532	–0.823	–0.805	2.405	0.508	4.524	275.977	3.224	4.346	6.484
16	*tert*–Butylamine	1.368	–0.067	–0.471	0.976	0.688	2.134	29.789	1.763	2.505	3.774
17	Iodoethane	1.583	–0.442	0.107	0.885	0.756	1.151	3.726	0.861	1.392	2.042
18	Methyl propionate	1.863	0.008	–0.639	1.291	0.615	2.769	85.508	2.270	3.093	4.799
19	2,2,3–Trimethyl–3–pentanol	2.701	–0.526	–0.656	1.760	0.508	3.315	81.074	2.521	3.516	5.311
20	1,2,3,5–Tetrachlorobenzene	2.983	–0.443	–0.621	2.002	0.647	3.728	164.806	2.689	3.736	5.621
21	2–Methyl–2–hexanol	2.711	–0.700	–0.693	1.738	0.438	3.281	62.828	2.483	3.408	5.169
22	1,2,4–Trimethylbenzene	2.685	–0.357	–0.627	1.856	0.691	3.566	150.596	2.617	3.606	5.401
23	3–Hexanol	2.530	–0.615	–0.647	1.609	0.462	3.033	48.313	2.339	3.204	4.856
24	2,2',4,4',6,6'–PCB	4.140	–1.014	–0.817	2.718	0.445	4.927	360.247	3.439	4.728	7.073
25	Ethanal	0.426	0.813	–0.368	0.410	0.742	1.756	41.334	1.431	1.903	3.005
26	1–Bromoheptane	3.285	–1.170	–0.647	1.988	0.360	3.459	40.610	2.414	3.348	4.967
27	Trifluoromethane	0.331	0.623	–0.772	0.404	0.500	2.033	66.528	2.057	2.636	4.117
28	N–methylaniline	2.495	–0.281	–0.625	1.748	0.717	3.430	135.935	2.566	3.487	5.217
29	Ethyl acetate	2.118	–0.059	–0.619	1.387	0.586	2.808	81.802	2.297	3.107	4.826
30	n–Heptanol	2.890	–1.081	–0.741	1.847	0.359	3.461	50.102	2.593	3.481	5.267
31	2–Hexanone	2.495	–0.426	–0.666	1.621	0.503	3.154	85.902	2.420	3.272	5.027
32	Ethylamine	0.602	0.125	–0.342	0.437	0.642	1.301	9.556	1.204	1.681	2.489
33	Bromobenzene	2.433	–0.148	–0.481	1.655	0.839	3.157	112.257	2.344	3.209	4.796
34	Hexanoic acid	2.759	–0.579	–0.760	1.775	0.432	3.466	114.248	2.705	3.597	5.534
35	1,2,3–Trimethylbenzene	2.662	–0.320	–0.615	1.847	0.706	3.549	152.368	2.609	3.595	5.378
36	2–Octanone	3.185	–1.074	–0.723	2.039	0.380	3.734	72.802	2.708	3.659	5.573
37	Cyclohexanol	2.280	–0.494	–0.575	1.553	0.593	2.995	57.886	2.348	3.180	4.743
38	Methanol	–0.699	0.699	–0.350	–0.350	0.408	0.778	5.000	0.778	1.079	1.716
39	2–Undecanone	3.761	–1.372	–0.898	2.389	0.275	4.398	129.494	3.117	4.191	6.361
40	Phenyl benzoate	3.789	–0.881	–0.881	2.557	0.455	4.806	370.608	3.440	4.629	6.953
41	2–Phenylpropanoic acid	3.053	–0.572	–0.755	2.096	0.562	4.058	231.823	3.054	4.085	6.162
42	2–Methyl–1–propanol	1.569	–0.272	–0.523	1.074	0.597	2.270	28.414	1.945	2.636	4.033
43	Heptanal	2.875	–0.885	–0.737	1.859	0.399	3.548	80.891	2.645	3.526	5.387
44	*cis*–1,2–Dichloroethene	1.427	0.113	–0.346	0.977	0.792	2.009	27.309	1.509	2.169	3.295
45	n–Propanol	1.265	–0.125	–0.476	0.846	0.591	1.903	19.965	1.716	2.301	3.549
46	2,2–Diphenylacetic acid	3.803	–0.921	–0.836	2.577	0.482	4.816	363.758	3.479	4.683	7.006
47	Hexadecanoic acid	4.644	–1.955	–1.083	2.924	0.179	5.311	206.413	3.700	4.934	7.487
48	Ethyl decanoate	4.097	–1.408	–0.972	2.570	0.246	4.698	173.317	3.331	4.489	6.839
49	Pentachlorobenzene	3.630	–0.677	–0.447	2.288	0.612	3.873	138.122	2.717	3.806	5.744
50	Ethylbenzene	2.616	–0.323	–0.592	1.788	0.718	3.429	125.935	2.526	3.450	5.144
51	3–Penten–2–ol	2.132	–0.347	–0.548	1.428	0.614	2.772	47.723	2.215	3.010	4.585
52	1–Methylnaphthalene	3.050	–0.467	–0.661	2.126	0.696	4.026	231.163	2.924	3.979	5.912
53	1,2–Dichloropropane	1.885	–0.453	–0.407	1.219	0.608	2.247	18.864	1.670	2.424	3.606
54	n–Heptane	2.714	–0.993	–0.645	1.723	0.404	3.152	34.075	2.260	3.128	4.616
55	Pentylamine	2.255	–0.748	–0.615	1.459	0.459	2.799	30.335	2.146	2.924	4.346
56	2–Methyl–2–pentanol	2.362	–0.455	–0.626	1.526	0.511	2.942	54.739	2.297	3.168	4.817
57	2–Phenylacetophenone	3.615	–0.778	–0.806	2.462	0.526	4.622	331.551	3.316	4.469	6.685
58	3–Methyl–1–butanol	2.036	–0.557	–0.599	1.358	0.511	2.701	35.820	2.188	2.952	4.496
59	Propylamine	1.301	–0.234	–0.453	0.873	0.602	1.881	15.964	1.602	2.204	3.272
60	n–Hexanol	2.584	–0.902	–0.688	1.660	0.404	3.155	42.075	2.425	3.253	4.936
61	2–Fluoroaniline	2.295	–0.135	–0.641	1.659	0.761	3.371	165.245	2.659	3.556	5.365
62	1–Penten–3–ol	2.080	–0.258	–0.579	1.392	0.615	2.777	55.690	2.204	2.997	4.542

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
63	Tribromomethane	1.771	–0.817	–0.052	1.124	0.756	1.675	5.814	1.108	1.831	2.888
64	Methyl cyclopropyl ether	2.146	–0.211	–0.595	1.415	0.594	2.799	59.827	2.176	2.997	4.587
65	Methyl butyrate	2.384	–0.196	–0.709	1.544	0.529	3.109	107.489	2.470	3.342	5.166
66	*o*–Xylene	2.447	–0.233	–0.578	1.727	0.766	3.365	133.245	2.494	3.422	5.110
67	Pentachloroethane	2.382	–0.554	–0.475	1.562	0.576	2.820	36.529	2.022	2.943	4.456
68	2–Phenyl–1–propanol	3.038	–0.701	–0.738	2.062	0.522	3.940	152.362	2.913	3.902	5.860
69	Pentamethylbenzene	3.039	–0.494	–0.680	2.058	0.610	3.876	190.006	2.809	3.894	5.839
70	Butylamine	1.830	–0.516	–0.541	1.199	0.527	2.380	22.937	1.903	2.602	3.866
71	Acetone	0.942	0.572	–0.439	0.737	0.779	2.065	60.704	1.716	2.335	3.657
72	Ethane	–0.301	0.903	0.000	0.000	0.894	1.000	8.000	0.602	0.903	1.204
73	1,2–Dichloroethane	1.589	–0.438	–0.326	1.015	0.621	1.850	10.725	1.385	2.032	2.969
74	Tetrachloroethene	2.071	–0.099	–0.461	1.391	0.696	2.652	53.966	1.936	2.795	4.295
75	Dichloromethane	0.939	–0.337	–0.158	0.620	0.736	1.169	4.316	0.911	1.432	2.037
76	*m*–Toluic acid	2.799	–0.391	–0.723	1.953	0.640	3.844	228.431	2.943	3.937	5.948
77	1–Iodopropane	2.196	–0.468	–0.117	1.239	0.623	1.846	10.134	1.352	2.007	2.972
78	Chlorobenzene	2.250	–0.141	–0.525	1.607	0.839	3.157	114.001	2.357	3.219	4.799
79	Trimethylamine	1.051	–0.097	–0.412	0.764	0.688	1.756	15.528	1.431	2.100	3.203
80	2–Bromobenzoic acid	2.971	–0.361	–0.663	1.992	0.657	3.825	226.150	2.922	3.917	5.923
81	Benzyl alcohol	2.434	–0.302	–0.629	1.732	0.718	3.430	133.935	2.627	3.522	5.279
82	Cyclododecanone	3.663	–1.201	–0.818	2.391	0.347	4.360	148.938	3.121	4.228	6.316
83	Octylbenzene	4.026	–1.358	–0.897	2.593	0.314	4.736	183.143	3.278	4.434	6.631
84	n–Butylbenzene	3.197	–0.715	–0.712	2.111	0.522	3.940	144.362	2.821	3.840	5.738
85	*m*–Xylene	2.462	–0.260	–0.587	1.733	0.754	3.374	132.396	2.500	3.429	5.125
86	Tertbutylbenzene	2.921	–0.537	–0.664	2.000	0.611	3.794	152.511	2.750	3.773	5.631
87	Ethyl butyrate	2.795	–0.374	–0.751	1.766	0.468	3.399	119.629	2.615	3.556	5.479
88	1,3,5–Trichlorobenzene	2.781	–0.377	–0.599	1.889	0.690	3.565	145.889	2.594	3.588	5.391
89	Tetrafluoromethane	0.602	0.602	–0.845	0.602	0.535	2.350	117.789	2.326	3.010	4.653
90	n–Octanol	3.157	–1.237	–0.788	2.010	0.323	3.730	58.374	2.742	3.681	5.559
91	1,6–Heptadiyne	2.474	–0.465	–0.651	1.736	0.616	3.453	120.180	2.519	3.350	5.066
92	1,4–Dimethylnaphthalene	3.208	–0.543	–0.688	2.211	0.651	4.158	249.906	3.004	4.103	6.104
93	2,3,4,5,6–PCB	3.999	–0.940	–0.793	2.649	0.472	4.830	345.940	3.384	4.640	6.933
94	2–Butyne	1.176	0.523	–0.407	0.898	0.908	2.182	62.373	1.644	2.301	3.549
95	Fluorobenzene	2.110	–0.050	–0.577	1.549	0.833	3.165	139.241	2.508	3.339	5.052
96	Trichloromethane	1.847	–0.696	–0.033	1.144	0.750	1.681	6.551	1.380	1.960	2.943
97	1–Bromobutane	2.287	–0.645	–0.391	1.383	0.532	2.370	17.445	1.721	2.463	3.640
98	Ethyl propyl ether	2.311	–0.476	–0.655	1.457	0.452	2.807	42.335	2.164	2.991	4.630
99	*o*–Methylaniline	2.387	–0.207	–0.602	1.701	0.765	3.366	141.245	2.542	3.459	5.157
100	1,1–Dichloroethane	1.466	–0.499	–0.206	0.970	0.737	1.697	8.288	1.253	1.897	2.699
101	2–Pentanone	2.114	–0.136	–0.598	1.390	0.594	2.805	77.802	2.220	3.010	4.645
102	Methyl decanoate	3.875	–1.299	–0.958	2.451	0.260	4.549	176.295	3.252	4.372	6.672
103	2–Pentanol	2.143	–0.462	–0.599	1.386	0.508	2.704	39.820	2.164	2.952	4.496
104	Cyclopropylamine	0.985	0.267	–0.350	0.787	0.899	1.903	29.539	1.602	2.204	3.235
105	Cyclopentene	1.580	0.024	–0.383	1.173	0.923	2.405	47.338	1.857	2.556	3.732
106	Ethylmethylamine	1.301	–0.125	–0.475	0.865	0.589	1.892	17.965	1.556	2.204	3.360
107	1,1,2,2–Tetrachloroethane	2.288	–0.674	–0.386	1.477	0.586	2.568	22.136	1.815	2.653	3.938
108	Methyl acetate	1.373	0.345	–0.626	0.994	0.666	2.456	90.016	2.083	2.812	4.386
109	Butyl pentanoate	3.555	–0.894	–0.876	2.231	0.333	4.124	145.519	3.000	4.079	6.242
110	Dibromomethane	1.373	–0.771	0.059	0.837	0.784	1.115	2.572	0.742	1.302	1.958
111	N,N–dimethylbenzylamine	2.978	–0.641	–0.729	2.031	0.549	3.889	153.750	2.829	3.849	5.781
112	2–Phenylethanol	2.754	–0.506	–0.684	1.905	0.610	3.694	143.021	2.777	3.721	5.584
113	1,2,4,5–Tetrachlorobenzene	2.996	–0.451	–0.625	2.007	0.641	3.735	164.898	2.692	3.740	5.630
114	2–Hexanol	2.521	–0.715	–0.666	1.616	0.439	3.074	47.593	2.369	3.221	4.889
115	Ethyl isobutyrate	2.665	–0.292	–0.728	1.708	0.504	3.326	123.051	2.582	3.526	5.425
116	Methanal	–0.778	1.380	–0.239	–0.239	0.555	1.415	24.000	0.903	1.204	1.903
117	3,3–Dimethyl–2–butanol	2.128	–0.372	–0.593	1.433	0.570	2.831	54.109	2.260	3.128	4.751
118	1,4–Dibromobenzene	2.848	–0.290	–0.505	1.852	0.745	3.385	126.597	2.463	3.406	5.129
119	1–Octene	2.948	–1.039	–0.705	1.908	0.399	3.545	56.891	2.519	3.429	5.099
120	n–Propyl–n–butylamine	2.980	–1.002	–0.746	1.876	0.359	3.461	50.102	2.494	3.429	5.148
121	Iodomethane	1.071	–1.071	0.535	0.535	0.960	0.035	0.085	0.035	0.337	0.304
122	Pentylbenzene	3.437	–0.897	–0.765	2.249	0.451	4.166	153.879	2.950	4.008	5.992
123	2–Methyl–3–pentanone	2.364	–0.158	–0.634	1.542	0.569	3.033	102.996	2.344	3.219	4.933
124	1,1,1–Trichloroethane	1.573	–0.358	–0.361	1.086	0.718	2.094	18.069	1.574	2.355	3.592

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
125	Benzyl acetate	3.171	–0.684	–0.813	2.153	0.497	4.166	230.462	3.078	4.121	6.245
126	1–Pentene	1.880	–0.358	–0.492	1.270	0.639	2.502	32.782	1.887	2.602	3.866
127	1–Chlorobutane	2.028	–0.622	–0.468	1.291	0.531	2.372	18.317	1.755	2.488	3.653
128	3–Methyl–2–butanol	1.942	–0.305	–0.562	1.297	0.571	2.591	42.236	2.114	2.910	4.428
129	*cis*–2–Butene	1.291	0.138	–0.395	0.920	0.785	2.017	29.169	1.556	2.204	3.318
130	3–Phenylpropionic acid	3.143	–0.718	–0.799	2.150	0.497	4.166	226.462	3.112	4.133	6.253
131	*p*–Xylene	2.477	–0.272	–0.592	1.739	0.746	3.383	132.502	2.504	3.432	5.132
132	4–Methyl–2–pentanol	2.351	–0.610	–0.642	1.544	0.482	2.982	49.065	2.330	3.186	4.834
133	n–Propyl isopropyl ether	2.543	–0.544	–0.689	1.618	0.435	3.077	55.593	2.322	3.221	4.956
134	Hexanal	2.568	–0.673	–0.683	1.674	0.456	3.253	72.601	2.481	3.305	5.066
135	Di–*tert*–butyl ketone	2.844	–0.493	–0.702	1.889	0.515	3.586	129.659	2.685	3.727	5.660
136	Ethyl methacrylate	2.615	–0.234	–0.728	1.711	0.562	3.383	143.498	2.619	3.561	5.465
137	Acetophenone	2.679	–0.315	–0.661	1.878	0.697	3.675	190.921	2.766	3.723	5.606
138	1–Fluorobutane	1.785	–0.378	–0.567	1.165	0.522	2.395	30.937	2.079	2.748	4.316
139	*sec*–Butyl acetate	2.641	–0.358	–0.736	1.709	0.485	3.357	124.342	2.602	3.533	5.436
140	1–Heptyne	2.611	–0.654	–0.648	1.730	0.510	3.328	77.128	2.408	3.253	4.897
141	Tetrachloromethane	1.608	–0.404	–0.342	1.105	0.721	2.090	16.829	1.548	2.335	3.579
142	1–Iodoheptane	3.637	–1.173	–0.574	2.089	0.360	3.459	40.272	2.406	3.342	4.966
143	Methyl acrylate	1.862	0.130	–0.654	1.317	0.692	2.865	115.576	2.330	3.135	4.835
144	Acetic acid	0.750	0.705	–0.542	0.634	0.710	2.134	80.704	1.887	2.486	3.874
145	1,8–Nonadiyne	3.064	–0.897	–0.760	2.070	0.446	3.983	138.449	2.825	3.760	5.672
146	Tetradecanoic acid	4.377	–1.772	–1.035	2.762	0.202	5.049	187.008	3.555	4.738	7.202
147	Dimethyl ether	0.301	0.301	–0.477	0.301	0.577	1.380	13.556	1.255	1.806	2.922
148	*p*–Trifluoromethylaniline	2.823	–0.471	–0.818	1.978	0.553	3.974	258.256	3.160	4.182	6.348
149	Methyl iso–propyl ether	1.572	–0.065	–0.551	1.064	0.578	2.288	36.414	1.881	2.636	4.113
150	Fluoromethane	–0.845	0.845	–0.423	–0.423	0.354	0.903	7.000	0.903	1.204	2.000
151	Methyl methacrylate	2.113	–0.025	–0.679	1.473	0.656	3.086	131.026	2.468	3.344	5.142
152	3–Methyl–3–pentanol	2.360	–0.361	–0.597	1.516	0.540	2.892	56.414	2.270	3.149	4.779
153	2,3,4,5–PCB	3.904	–0.914	–0.790	2.595	0.482	4.753	326.450	3.339	4.567	6.824
154	Methyl 2–phenylacetate	3.118	–0.640	–0.795	2.130	0.521	4.123	227.825	3.062	4.106	6.215
155	3–Methyl–2–pentanone	2.341	–0.245	–0.611	1.540	0.583	3.022	89.926	2.356	3.219	4.933
156	Ethyl formate	1.664	0.196	–0.679	1.098	0.583	2.580	94.782	2.170	2.881	4.492
157	3–Methylpentane	2.214	–0.579	–0.524	1.429	0.536	2.659	28.622	1.973	2.792	4.127
158	4–Bromobenzoic acid	3.085	–0.420	–0.693	2.022	0.626	3.864	225.745	2.943	3.940	5.970
159	Pentanal	2.210	–0.419	–0.622	1.458	0.530	2.917	64.552	2.292	3.049	4.695
160	2–Methyl–4–penten–3–ol	2.298	–0.397	–0.617	1.548	0.577	3.035	66.060	2.356	3.219	4.864
161	1,1,2–Trichloroethane	1.918	–0.483	–0.387	1.238	0.609	2.244	17.934	1.646	2.406	3.591
162	*iso*–Butylamine	1.609	–0.336	–0.505	1.098	0.601	2.260	24.414	1.845	2.556	3.812
163	Dibutylamine	3.243	–1.159	–0.794	2.038	0.323	3.730	58.374	2.651	3.636	5.450
164	Triphenylamine	4.178	–1.071	–0.866	2.813	0.449	5.168	439.873	3.628	4.917	7.304
165	N–ethylpiperidine	2.686	–0.619	–0.629	1.762	0.523	3.291	66.900	2.431	3.363	5.024
166	Dimethylamine	0.477	0.125	–0.389	0.389	0.633	1.301	9.556	1.146	1.681	2.598
167	Methyl benzoate	2.861	–0.409	–0.740	1.983	0.617	3.884	221.922	2.930	3.940	5.959
168	1,3–Dibromobenzene	2.808	–0.279	–0.494	1.840	0.755	3.373	126.468	2.459	3.399	5.113
169	*cis*–1,3–Dichloropropene	2.004	–0.207	–0.455	1.299	0.661	2.468	36.010	1.825	2.574	3.885
170	Butyric acid	2.041	–0.027	–0.653	1.336	0.585	2.818	97.802	2.354	3.128	4.839
171	2–Aminonaphthalene	3.020	–0.486	–0.688	2.117	0.677	4.049	238.561	2.969	4.016	5.968
172	n–Propylbenzene	2.926	–0.522	–0.654	1.958	0.611	3.693	135.021	2.679	3.654	5.456
173	*m*–Methylaniline	2.395	–0.235	–0.610	1.705	0.753	3.376	140.396	2.549	3.467	5.172
174	1–Pentyne	1.834	–0.082	–0.495	1.272	0.738	2.617	60.627	1.973	2.681	4.055
175	2,2,2–Trifluoroethanol	1.343	0.223	–0.806	0.997	0.551	2.629	110.998	2.491	3.253	5.019
176	5–Hexyne–2–one	2.332	–0.292	–0.669	1.629	0.615	3.315	128.857	2.522	3.360	5.142
177	Methyl 3–phenylpropionate	3.363	–0.839	–0.844	2.269	0.445	4.344	237.876	3.182	4.259	6.449
178	2–Propanol	1.041	0.146	–0.438	0.738	0.654	1.820	24.528	1.644	2.255	3.481
179	1,1,2–Trichlorotrifluoroethane	2.250	–0.129	–0.690	1.494	0.547	3.068	124.561	2.656	3.561	5.497
180	1–Iodobutane	2.670	–0.655	–0.277	1.514	0.532	2.370	17.107	1.706	2.452	3.638
181	2,2',3,3',4,4'–PCB	4.167	–1.038	–0.828	2.731	0.433	4.948	364.823	3.449	4.739	7.094
182	Bromocyclohexane	2.573	–0.646	–0.472	1.667	0.597	2.989	43.663	2.182	3.047	4.474
183	Ethyne	–0.477	1.431	0.000	0.000	0.949	1.477	27.000	0.778	1.079	1.556
184	2–Cyclohexen–1–one	2.212	–0.135	–0.571	1.562	0.746	3.123	119.183	2.433	3.267	4.924
185	Azobenzene	3.716	–0.854	–0.865	2.512	0.465	4.726	343.883	3.359	4.518	6.775
186	1,2–Dimethylcyclohexane	2.593	–0.672	–0.573	1.737	0.562	3.219	57.387	2.358	3.283	4.821

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
187	Isopropyl benzoate	3.372	–0.654	–0.814	2.243	0.494	4.255	250.393	3.119	4.223	6.383
188	*p*–Methylaniline	2.409	–0.250	–0.616	1.711	0.743	3.387	140.525	2.556	3.474	5.188
189	2,4–Dimethyl–3–pentanol	2.599	–0.607	–0.665	1.691	0.475	3.200	61.294	2.443	3.379	5.109
190	1,2,3,5–Tetramethylbenzene	2.882	–0.448	–0.662	1.968	0.636	3.742	169.949	2.724	3.761	5.639
191	Hexachlorobenzene	3.322	–0.568	–0.657	2.196	0.580	4.010	202.874	2.855	3.991	6.012
192	2,4–Dimethyl–2–pentanol	2.546	–0.616	–0.671	1.671	0.476	3.195	64.281	2.447	3.376	5.117
193	2–Methyl–2–propanol	1.350	0.058	–0.487	0.960	0.675	2.158	37.789	1.857	2.584	3.970
194	*o*–Toluic acid	2.778	–0.355	–0.710	1.943	0.657	3.825	229.114	2.932	3.926	5.925
195	3–Pentanol	2.156	–0.377	–0.580	1.381	0.528	2.670	40.621	2.140	2.939	4.470
196	Hexylamine	2.609	–0.945	–0.677	1.675	0.404	3.154	38.075	2.350	3.195	4.750
197	1,3,5–Trimethylbenzene	2.681	–0.367	–0.628	1.855	0.690	3.565	149.609	2.617	3.606	5.400
198	3–Hexanone	2.520	–0.304	–0.642	1.616	0.535	3.102	87.489	2.386	3.255	4.995
199	Phenanthrene	3.519	–0.690	–0.732	2.428	0.617	4.532	331.403	3.243	4.397	6.521
200	2,3–Ddimethyl–2–butanol	2.168	–0.316	–0.590	1.445	0.568	2.834	58.109	2.250	3.128	4.751
201	Ethyl heptanoate	3.544	–0.959	–0.873	2.228	0.327	4.136	145.692	3.018	4.084	6.249
202	1,2–Dimethylnaphthalene	3.221	–0.553	–0.693	2.218	0.642	4.171	251.120	3.010	4.108	6.114
203	2,2',3,3',5,5',6,6'–PCB	4.306	–1.079	–0.827	2.817	0.424	5.066	399.725	3.521	4.860	7.267
204	1,3–Cyclohexadiene	1.965	–0.123	–0.480	1.429	0.850	2.851	76.268	2.158	2.937	4.330
205	n–Propyl–sec–butylamine	2.803	–0.756	–0.666	1.780	0.439	3.290	53.854	2.407	3.344	5.005
206	1–Fluoropentane	2.215	–0.644	–0.635	1.431	0.458	2.804	38.335	2.301	3.049	4.751
207	2,2–Dimethyl–3–pentanol	2.541	–0.543	–0.645	1.653	0.507	3.145	62.886	2.422	3.356	5.078
208	2–Methylnaphthalene	3.079	–0.498	–0.674	2.138	0.677	4.048	230.561	2.935	3.991	5.938
209	Methylcyclohexane	2.363	–0.618	–0.535	1.598	0.596	2.990	45.886	2.212	3.070	4.487
210	Benzylamine	2.448	–0.307	–0.625	1.738	0.718	3.430	131.935	2.604	3.505	5.233
211	Ethyl Nonanoate	3.928	–1.272	–0.941	2.466	0.268	4.526	163.971	3.235	4.364	6.658
212	2–Heptanone	2.820	–0.679	–0.728	1.818	0.429	3.468	94.304	2.596	3.502	5.362
213	Cyclopentane	1.653	–0.255	–0.380	1.176	0.791	2.301	27.782	1.778	2.477	3.556
214	*trans*–1,2–Dichloroethene	1.427	0.113	–0.346	0.977	0.792	2.009	27.309	1.509	2.169	3.295
215	Ethyl propionate	2.472	–0.159	–0.706	1.569	0.527	3.111	111.489	2.458	3.342	5.166
216	Anthracene	3.552	–0.696	–0.727	2.454	0.628	4.573	349.080	3.257	4.409	6.547
217	Ethyl benzoate	3.229	–0.548	–0.780	2.143	0.542	4.093	235.307	3.034	4.096	6.196
218	*p*–Ethyltoluene	2.854	–0.442	–0.643	1.917	0.646	3.633	143.352	2.649	3.636	5.443
219	Diethylamine	1.906	–0.358	–0.556	1.215	0.519	2.387	26.937	1.857	2.602	3.939
220	Chloroethylene	0.690	0.370	–0.210	0.569	0.929	1.506	16.404	1.148	1.658	2.472
221	*sec*–Butylamine	1.697	–0.270	–0.502	1.120	0.596	2.265	26.414	1.833	2.556	3.812
222	1–Chloroheptane	3.065	–1.161	–0.692	1.922	0.360	3.459	41.482	2.434	3.362	4.974
223	*p*–Cymene	3.014	–0.562	–0.681	2.028	0.589	3.819	155.889	2.762	3.795	5.680
224	Isopropylbenzene	2.791	–0.450	–0.633	1.909	0.651	3.631	138.310	2.649	3.626	5.407
225	2–Nonanone	3.341	–1.067	–0.820	2.134	0.336	3.980	111.539	2.882	3.880	5.910
226	2,2–Dimethyl–1–propanol	1.783	–0.327	–0.558	1.235	0.606	2.549	38.998	2.114	2.894	4.417
227	Diiodomethane	1.849	–1.247	0.297	1.075	0.805	1.091	1.896	0.654	1.240	1.946
228	Methyl *tert*–butyl ether	1.799	–0.118	–0.577	1.231	0.592	2.564	50.998	2.057	2.894	4.484
229	Formic acid	–0.125	1.079	–0.602	0.176	0.548	1.954	74.335	1.644	2.107	3.287
230	2,3–Dimethyl–2–pentanol	2.557	–0.512	–0.639	1.659	0.506	3.146	66.886	2.422	3.356	5.078
231	Ethyl acrylate	2.401	–0.115	–0.703	1.580	0.590	3.187	127.858	2.502	3.380	5.200
232	Hexamethylbenzene	3.190	–0.554	–0.702	2.145	0.579	4.011	210.314	2.892	4.019	6.025
233	2–Aminooctane	3.127	–1.165	–0.768	1.990	0.342	3.674	57.963	2.647	3.613	5.382
234	5–Methyl–2–octanone	3.224	–0.947	–0.781	2.073	0.378	3.880	113.524	2.835	3.837	5.834
235	n–Pentane	1.959	–0.606	–0.490	1.263	0.531	2.373	18.937	1.778	2.505	3.670
236	Hexachloroethane	2.549	–0.579	–0.508	1.683	0.563	3.038	47.785	2.165	3.149	4.782
237	3–Methyl–2–butanone	1.909	–0.001	–0.559	1.304	0.662	2.702	81.123	2.170	2.968	4.571
238	Iodobenzene	2.752	–0.151	–0.409	1.733	0.839	3.157	111.581	2.339	3.205	4.796
239	1–Aminonaphthalene	3.004	–0.453	–0.677	2.107	0.695	4.026	239.163	2.955	4.004	5.942
240	Methyl butyl ether	2.213	–0.586	–0.654	1.426	0.456	2.804	38.335	2.188	2.991	4.630
241	*o*–Ethyltoluene	2.802	–0.392	–0.622	1.897	0.673	3.601	143.931	2.634	3.616	5.400
242	2,2',3,3',4,4',6–PCB	4.254	–1.070	–0.833	2.781	0.423	5.019	382.514	3.491	4.806	7.194
243	Iso–propylamine	1.067	0.000	–0.415	0.761	0.676	1.778	18.528	1.532	2.158	3.225
244	Methyl sec–butyl ether	2.071	–0.311	–0.602	1.347	0.525	2.674	44.621	2.114	2.939	4.542
245	2–Methylpentane	2.217	–0.657	–0.546	1.438	0.514	2.694	27.820	1.991	2.806	4.154
246	*trans*–Stilbene	3.740	–0.977	–0.833	2.532	0.466	4.726	295.883	3.330	4.480	6.686
247	Cyclohexanone	2.265	–0.239	–0.572	1.561	0.665	3.070	97.062	2.395	3.225	4.868
248	1,3,5–Cycloheptatriene	2.246	–0.234	–0.537	1.621	0.810	3.183	105.584	2.380	3.225	4.767

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
249	3–Chloroaniline	2.425	–0.238	–0.603	1.714	0.753	3.376	139.508	2.544	3.463	5.170
250	1,2–Dibromobenzene	2.826	–0.308	–0.456	1.853	0.769	3.362	117.059	2.437	3.370	5.051
251	Ethyl–iso–propylamine	2.174	–0.439	–0.597	1.401	0.506	2.702	37.820	2.057	2.885	4.358
252	Chloromethane	0.161	–0.161	0.081	0.081	0.769	0.228	0.690	0.228	0.529	0.470
253	Trichloroethylene	1.779	–0.010	–0.409	1.201	0.748	2.353	39.481	1.742	2.514	3.853
254	2,2',3,3',4,5,5',6,6'–PCB	4.408	–1.126	–0.841	2.873	0.407	5.148	419.969	3.568	4.930	7.376
255	1,2,3,4–tetramethylbenzene	2.862	–0.415	–0.651	1.958	0.652	3.722	170.666	2.714	3.753	5.624
256	n–Octane	3.010	–1.154	–0.705	1.903	0.360	3.459	42.102	2.447	3.371	4.983
257	1–Octyne	2.913	–0.868	–0.707	1.909	0.439	3.618	85.681	2.580	3.481	5.231
258	Benzyl benzoate	3.966	–1.049	–0.926	2.657	0.399	4.974	373.644	3.531	4.745	7.133
259	Bromochloromethane	1.208	–0.502	–0.036	0.742	0.760	1.143	3.444	0.835	1.372	2.000
260	2–Butanone	1.644	0.205	–0.517	1.110	0.697	2.425	70.016	1.987	2.702	4.195
261	2–Methylbutane	1.768	–0.412	–0.452	1.168	0.605	2.250	20.414	1.716	2.459	3.636
262	1–Hexene	2.299	–0.629	–0.576	1.520	0.535	2.904	40.552	2.134	2.924	4.346
263	2,6–PCB	3.612	–0.771	–0.749	2.442	0.547	4.525	289.230	3.212	4.374	6.519
264	*o*–Chlorotoluene	2.483	–0.238	–0.567	1.739	0.766	3.365	132.006	2.486	3.415	5.107
265	Bromomethane	0.595	–0.595	0.298	0.298	0.893	0.098	0.254	0.098	0.399	0.328
266	Dibutyl ether	3.234	–1.054	–0.811	2.027	0.322	3.730	66.374	2.688	3.681	5.619
267	Dodecanoic acid	4.072	–1.557	–0.980	2.577	0.233	4.748	168.006	3.391	4.517	6.881
268	*m*–Chlorotoluene	2.501	–0.264	–0.576	1.746	0.754	3.374	131.156	2.492	3.423	5.122
269	2–PCB	3.479	–0.721	–0.739	2.371	0.570	4.424	271.704	3.154	4.280	6.374
270	5–Nonanone	3.358	–0.864	–0.794	2.122	0.371	3.903	112.686	2.827	3.854	5.862
271	Dimethyl–*n*–butylamine	2.540	–0.799	–0.653	1.634	0.440	3.071	39.593	2.250	3.131	4.699
272	2,5–PCB	3.641	–0.797	–0.761	2.453	0.534	4.544	288.646	3.221	4.385	6.540
273	Tripropylamine	3.321	–0.995	–0.757	2.092	0.364	3.786	73.172	2.697	3.742	5.620
274	Pentyl acetate	3.109	–0.802	–0.792	1.987	0.378	3.737	106.802	2.804	3.773	5.806
275	Benzene	1.909	0.000	–0.477	1.431	0.949	2.909	98.176	2.210	2.988	4.419
276	Vinyl ecetate	1.933	0.089	–0.671	1.348	0.656	2.913	119.948	2.346	3.149	4.861
277	3–Heptanone	2.785	–0.413	–0.667	1.778	0.506	3.337	95.114	2.489	3.441	5.248
278	Aniline	2.134	–0.105	–0.563	1.564	0.836	3.160	123.241	2.418	3.267	4.855
279	1,2,4–Trichlorobenzene	2.787	–0.360	–0.598	1.891	0.691	3.566	147.876	2.594	3.588	5.392
280	Benzoic acid	2.583	–0.234	–0.704	1.834	0.691	3.683	230.983	2.840	3.790	5.721
281	4–Phenylbutyric acid	3.383	–0.900	–0.844	2.284	0.430	4.378	236.376	3.225	4.281	6.480
282	Phenyl formate	2.651	–0.324	–0.729	1.871	0.645	3.756	219.860	2.873	3.818	5.784
283	2–Methyl–2–butanol	1.940	–0.184	–0.551	1.275	0.596	2.559	46.998	2.086	2.894	4.417
284	1–Hexene–3–ol	2.466	–0.511	–0.645	1.618	0.525	3.123	63.682	2.396	3.255	4.923
285	Benzophenone	3.807	–0.955	–0.856	2.571	0.455	4.805	338.608	3.415	4.596	6.882
286	2,2',3,3',6,6'–PCB	4.099	–0.984	–0.802	2.703	0.459	4.899	361.584	3.426	4.714	7.042
287	Azulene	2.860	–0.385	–0.625	2.025	0.759	3.867	211.262	2.830	3.836	5.685
288	Isobutene	1.051	0.204	–0.356	0.820	0.876	1.924	28.704	1.505	2.158	3.267
289	Propyl formate	2.120	–0.134	–0.722	1.379	0.515	2.937	102.552	2.386	3.167	4.918
290	Phenyl acetate	2.925	–0.447	–0.760	2.011	0.582	3.933	228.635	2.949	3.958	5.994
291	2–Bromopropane	1.473	–0.346	–0.226	0.950	0.723	1.714	10.290	1.307	1.976	2.968
292	2,4,6–PCB	3.774	–0.852	–0.773	2.524	0.510	4.648	306.418	3.280	4.477	6.684
293	1–Hexyne	2.258	–0.397	–0.579	1.521	0.607	2.997	68.770	2.210	2.991	4.511
294	Propyl acetate	2.446	–0.255	–0.734	1.575	0.497	3.161	109.902	2.487	3.360	5.199
295	Hexylbenzene	3.664	–1.121	–0.797	2.385	0.395	4.372	152.529	3.058	4.147	6.185
296	2,4–Dimethylpentane	2.434	–0.751	–0.599	1.593	0.485	2.977	37.065	2.176	3.062	4.550
297	3–Bromopropene	1.763	–0.048	–0.256	1.097	0.780	2.027	23.853	1.525	2.162	3.249
298	Diphenylcarbinol	3.617	–0.872	–0.811	2.458	0.499	4.608	287.427	3.300	4.451	6.637
299	Difluoromethane	–0.067	0.669	–0.661	0.117	0.447	1.602	29.556	1.663	2.107	3.351
300	Dichlorodifluoromethane	1.301	0.347	–0.544	0.903	0.626	2.241	68.013	2.097	2.796	4.388
301	Naphthalene	2.878	–0.395	–0.631	2.033	0.748	3.882	211.486	2.838	3.844	5.702
302	*p*–Toluic acid	2.825	–0.415	–0.732	1.963	0.626	3.865	228.709	2.954	3.948	5.972
303	2–Chloroaniline	2.413	–0.209	–0.595	1.709	0.765	3.366	140.357	2.537	3.455	5.155
304	n–Butanol	1.799	–0.441	–0.558	1.177	0.525	2.387	26.937	2.000	2.681	4.104
305	Butyl acetate	2.775	–0.507	–0.781	1.777	0.430	3.467	118.248	2.655	3.579	5.521
306	Methacrylic acid	1.709	0.182	–0.614	1.248	0.751	2.792	121.009	2.344	3.125	4.802
307	Propionic acid	1.531	0.404	–0.620	1.018	0.653	2.480	106.016	2.117	2.835	4.400
308	*trans*–Cinnamic acid	3.122	–0.654	–0.793	2.155	0.536	4.185	247.274	3.115	4.144	6.250
309	5–Hexene–2–one	2.394	–0.378	–0.667	1.627	0.560	3.240	100.380	2.474	3.318	5.077
310	5–Methyl–2–hexanone	2.669	–0.606	–0.707	1.757	0.465	3.395	95.725	2.564	3.471	5.311

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
311	1–Bromooctane	3.530	–1.311	–0.707	2.138	0.323	3.729	48.882	2.581	3.564	5.290
312	1,3–Dimethylnaphthalene	3.224	–0.565	–0.696	2.219	0.639	4.174	249.190	3.011	4.110	6.120
313	Octanoic acid	3.291	–0.986	–0.846	2.101	0.336	3.981	131.539	2.979	3.962	6.071
314	1–Bromopropane	1.801	–0.450	–0.255	1.091	0.622	1.848	10.473	1.372	2.022	2.974
315	n–Pentanol	2.227	–0.693	–0.628	1.441	0.459	2.801	34.335	2.230	2.991	4.555
316	1,2,4,5–Tetramethylbenzene	2.882	–0.439	–0.660	1.966	0.641	3.735	169.858	2.721	3.761	5.641
317	Butyl methacrylate	3.207	–0.647	–0.829	2.059	0.418	3.911	161.241	2.898	3.934	6.019
318	1,5–Hexadiene	2.213	–0.500	–0.577	1.526	0.610	2.991	54.770	2.210	2.991	4.463
319	2–Chloropropane	1.237	–0.269	–0.321	0.855	0.713	1.725	11.598	1.351	2.010	2.984
320	1,1–Dichloroethylene	1.165	0.157	–0.307	0.869	0.892	1.911	26.224	1.461	2.123	3.246
321	1–Bromopentane	2.679	–0.838	–0.494	1.619	0.461	2.795	24.843	1.995	2.811	4.164
322	Ethyl octanoate	3.745	–1.123	–0.908	2.352	0.295	4.339	154.758	3.131	4.229	6.462
323	2,4–Dimethyl–3–pentanone	2.593	–0.333	–0.658	1.700	0.541	3.257	102.566	2.484	3.425	5.230
324	4–PCB	3.537	–0.775	–0.761	2.392	0.545	4.463	270.490	3.172	4.301	6.419
325	Isobutylbenzene	3.074	–0.667	–0.699	2.067	0.549	3.888	145.750	2.797	3.816	5.698
326	Cyclohexane	2.085	–0.528	–0.477	1.431	0.651	2.708	35.451	2.033	2.812	4.067
327	Methyl *iso*–butyl ether	2.020	–0.457	–0.628	1.342	0.508	2.704	39.820	2.140	2.952	4.569
328	Methyl n–propyl ether	1.783	–0.305	–0.594	1.160	0.516	2.395	30.937	1.945	2.681	4.187
329	2–Methyl–3–pentanol	2.377	–0.489	–0.618	1.538	0.510	2.938	50.750	2.297	3.168	4.799
330	1,3–Dichlorobenzene	2.538	–0.268	–0.566	1.758	0.754	3.374	129.916	2.483	3.416	5.118
331	Chloroethane	0.866	–0.249	–0.196	0.582	0.722	1.187	4.936	0.958	1.470	2.081
332	Methylamine	–0.477	0.477	–0.239	–0.239	0.500	0.602	3.000	0.602	0.903	1.301
333	Salicylic acid	2.709	–0.316	–0.740	1.914	0.656	3.826	246.114	2.981	3.971	5.987
334	3,3–Dimethyl–1–butanol	2.209	–0.589	–0.628	1.488	0.515	2.938	46.739	2.330	3.168	4.817
335	1,2,3–Trichloropropane	2.358	–0.618	–0.469	1.491	0.537	2.657	26.451	1.917	2.750	4.092
336	Bromoethane	1.171	–0.379	–0.065	0.713	0.746	1.162	4.064	0.890	1.416	2.047
337	1,2–Dichlorobenzene	2.516	–0.242	–0.556	1.751	0.766	3.364	130.766	2.478	3.409	5.103
338	1,4–Dichlorobenzene	2.564	–0.280	–0.573	1.766	0.745	3.385	130.045	2.488	3.423	5.134
339	2,2',4',5–PCB	3.931	–0.927	–0.799	2.602	0.476	4.759	324.673	3.342	4.574	6.839
340	Heptylamine	2.912	–1.116	–0.732	1.861	0.359	3.460	46.102	2.526	3.429	5.099
341	2–Methyl–1–butanol	2.068	–0.479	–0.580	1.354	0.533	2.667	36.621	2.164	2.939	4.470
342	1–Chloropropane	1.524	–0.408	–0.354	0.984	0.618	1.854	11.345	1.417	2.056	2.994
343	3–PCB	3.504	–0.756	–0.751	2.380	0.556	4.443	270.116	3.164	4.291	6.396
344	2,2',4,5,5'–PCB	4.043	–0.979	–0.811	2.664	0.457	4.848	342.549	3.393	4.656	6.964
345	3–Ethyl–3–pentanol	2.682	–0.498	–0.631	1.707	0.500	3.167	65.984	2.422	3.363	5.079
346	Triethylamine	2.529	–0.513	–0.620	1.604	0.485	2.988	49.029	2.225	3.128	4.742
347	4–Penten–1–ol	2.124	–0.547	–0.629	1.449	0.533	2.909	48.552	2.292	3.049	4.630
348	Isopropyl acetate	2.271	–0.148	–0.705	1.505	0.540	3.079	115.310	2.449	3.325	5.136
349	2–Phenylacetic acid	2.875	–0.515	–0.750	2.002	0.584	3.933	216.635	2.987	3.970	6.002
350	Piperidine	2.005	–0.403	–0.522	1.387	0.646	2.714	43.451	2.100	2.879	4.243
351	Dipropyl ether	2.668	–0.690	–0.715	1.677	0.400	3.157	50.075	2.362	3.253	5.006
352	Trichlorofluoromethane	1.447	0.125	–0.447	1.000	0.672	2.174	43.125	1.910	2.630	4.151
353	3–Methyl–3–hexanol	2.701	–0.581	–0.659	1.721	0.473	3.219	64.351	2.447	3.382	5.120
354	4–Chloroaniline	2.442	–0.253	–0.609	1.721	0.744	3.387	139.637	2.551	3.470	5.186
355	3–Methyl–2–pentanol	2.366	–0.510	–0.613	1.534	0.511	2.938	50.750	2.305	3.168	4.799
356	Pentanoic Acid	2.431	–0.325	–0.709	1.574	0.500	3.160	105.902	2.542	3.379	5.212
357	1–Aminooctane	3.178	–1.267	–0.781	2.023	0.323	3.729	54.374	2.681	3.636	5.406
358	Dipropylamine	2.680	–0.816	–0.692	1.692	0.402	3.155	42.075	2.314	3.195	4.805
359	Propanal	1.239	0.292	–0.465	0.870	0.718	2.121	49.345	1.792	2.380	3.716
360	4–Methyl–2–pentanone	2.318	–0.340	–0.642	1.549	0.550	3.069	87.310	2.382	3.238	4.968
361	Diethyl ether	1.887	–0.202	–0.588	1.192	0.508	2.401	34.937	1.924	2.681	4.187
362	Ethanol	0.556	0.301	–0.379	0.399	0.604	1.380	13.556	1.342	1.806	2.820
363	2,3–Dimethyl–3–pentanol	2.556	–0.466	–0.630	1.656	0.517	3.128	67.750	2.408	3.349	5.065
364	Decanoic acid	3.716	–1.302	–0.918	2.360	0.275	4.398	149.494	3.202	4.262	6.510
365	Cyclohexylamine	2.295	–0.551	–0.565	1.564	0.595	2.993	51.886	2.286	3.128	4.590
366	1–Ethylnaphthalene	3.315	–0.585	–0.692	2.246	0.630	4.192	242.459	3.020	4.117	6.121
367	Methylcyclopentane	2.003	–0.383	–0.458	1.383	0.704	2.644	38.006	2.000	2.795	4.078
368	*trans*–2–Butene	1.291	0.138	–0.395	0.920	0.785	2.017	29.169	1.556	2.204	3.318
369	2–Hexene–4–ol	2.453	–0.438	–0.650	1.600	0.538	3.097	68.501	2.386	3.255	4.938
370	Cycloheptane	2.401	–0.711	–0.535	1.623	0.575	3.015	43.108	2.225	3.070	4.451
371	1–Nonyne	3.178	–1.053	–0.759	2.066	0.385	3.874	94.411	2.731	3.681	5.526
372	Butanal	1.780	–0.107	–0.551	1.197	0.623	2.534	56.782	2.068	2.748	4.256

Table 2. (Continued)

No.	Compound name	Sh1	Sh2	Sh3	Sh4	Sh5	Sh6	Sh7	Sh8	Sh9	Sh10
373	Methyl–n–butylamine	2.258	–0.684	–0.628	1.455	0.458	2.800	32.335	2.121	2.924	4.409
374	2–Decanone	3.561	–1.228	–0.861	2.267	0.303	4.199	120.441	3.005	4.042	6.146
375	1,2,3–Trichlorobenzene	2.753	–0.331	–0.585	1.880	0.706	3.549	148.648	2.586	3.578	5.370
376	*n*–Nonanol	3.395	–1.377	–0.831	2.155	0.294	3.969	66.862	2.875	3.861	5.821
377	3–Bromobenzoic acid	3.027	–0.396	–0.680	2.007	0.640	3.844	225.467	2.932	3.929	5.945
378	Methyl 4–phenylbutyrate	3.578	–1.006	–0.883	2.390	0.391	4.538	247.842	3.288	4.396	6.656
379	Ethyl hexanoate	3.322	–0.779	–0.835	2.091	0.366	3.913	136.792	2.895	3.924	6.017

2.5 Software

All calculations were run on a Pentium III personal computer with windows XP operating system. All the necessary programs for PCA, PCR, ANN and other statistical analysis were written in MATLAB (ver. 6.5, MathWork Inc.).

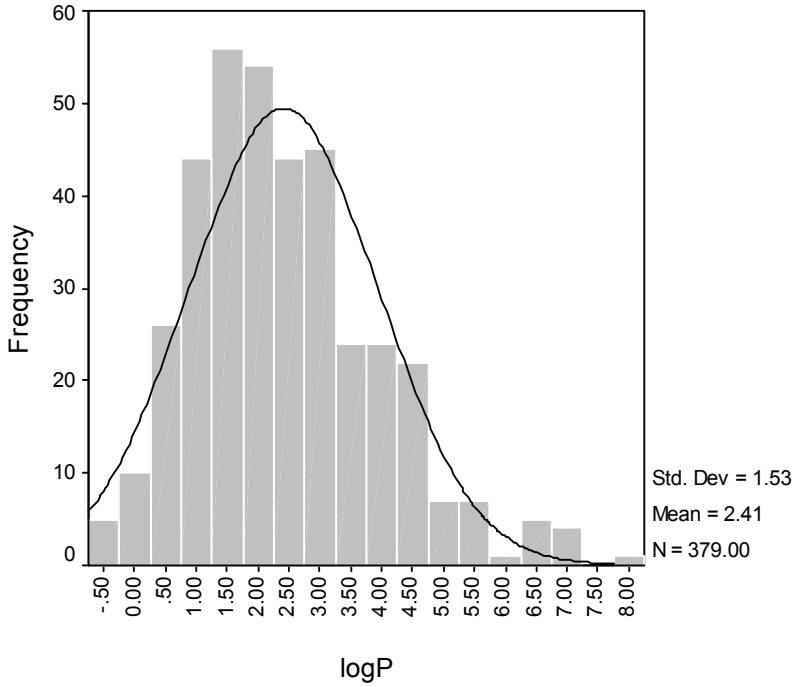

Figure 1. Histogram of the distribution of the experimental logP for the total data set of 379 organic compounds used in this study. The solid curve is the fitting of the logP data to the normal distribution.

3 RESULTS AND DISCUSSION

Table 1 lists the name of the compounds used in this study and their corresponding experimental logP value. In this list, the range of experimental logP values is (–0.66) – (8.16) log units with an average and standard deviation equal to 2.41 and 1.53, respectively. To show the distribution of the experimental logP, their histogram–plot is shown in Figure 1. The compounds range in size from 26 (molecular weight of ethyne) to 464.5 (molecular weight of 2,2',3,3',4,5,5',6,6'–PCB). The 10 Sh topological indices, which are easily calculated from the two–dimensional structure of the molecules, are presented in Table 2. Since there is some collinearity between the Sh indices;

orthogonal transformation of the indices was performed by PCA. PCR and PC–ANN methods were used to model the respective linear and nonlinear relationships between the extracted PCs and logP values.

3.1 Linear Modeling

As it is shown in Table 1, a wide variety of organic molecules including saturated and unsaturated hydrocarbons, halogenated hydrocarbons, polychlorinated biphenyls (PCBs), esters, aldehydes, organic acids, alcohols, ethers, amines, and aromatic compounds are included in the Table. Therefore, for each subset of molecules separate PCR models based on the eigenvalue ranking and correlation ranking were obtained. The results obtained by the correlation ranking procedure are shown in Table 3. As can be seen, the number of PCs, used in the QSPR model of each subset was different. The least number of factors (*i.e.* 3 factors) is used for modeling the logP of alkynes and aldehydes, while the highest number of factors is used by aromatics (8 factors) and halogenated hydrocarbons (7 factors).

Table 3. Linear multivariate regression models and statistical parameters of compounds properties using *f* indices. N denotes number of structures.

Subset	*N*	Equation	R^2	*SE*	*RMS*	*F*	R^2_{CV}
Alkanes	15	logP =2.85 + 0.47 F_1 – 0.15 F_2 + 0.04 F_7 – 0.04 F_6 + 0.03 F_4	0.9973	0.033	0.025	654	0.9909
Alkenes	13	logP = 2.85 + 0.99 F_1 – 0.39 F_2 + 0.30 F_7 – 0.16 F_3 – 0.13 F_5	0.9835	0.191	0.140	83	0.9270
Alkynes	9	logP =2.24 + 0.89 F_1 – 0.07 F_2 – 0.05 F_4	0.9990	0.036	0.027	1660	0.9945
Ethers	15	logP = 1.39 + 0.65 F_1 – 0.12 F_2 – 0.09 F_8 – 0.07 F_3 – 0.07 F_4	0.9940	0.066	0.051	297	0.9633
Alcohols	49	logP = 1.34 + 0.65 F_1 – 0.20 F_2 + 0.06 F_6 – 0.06 F_7 + 0.04 F_9 – 0.04 F_8	0.9721	0.125	0.115	244	0.9602
Aldehydes	7	logP = 1.11 + 0.71 F_1 – 0.24 F_2 + 0.17 F_3	0.9966	0.063	0.041	296	0.8461
Ketones	29	logP = 1.80 + 1.11 F_1 – 0.20 F_2 + 0.19 F_9 + 0.18 F_6 – 0.11 F_8	0.9693	0.229	0.204	145	0.9429
Amines	46	logP = 1.45 + 1.05 F_1 + 0.30 F_4 – 0.19 F_9 – 0.13 F_3 – 0.12 F_2	0.9835	0.154	0.144	478	0.9776
Esters	38	logP = 2.07 + 1.13 F_1 – 0.22 F_2 – 0.21 F_7 + 0.14 F_5 + 0.12 F_9 – 0.09 F_6	0.9807	0.182	0.165	262	0.9737
Acids	26	logP = 2.38 + 1.56 F_1 – 0.42 F_2 + 0.41 F_6 – 0.28 F_8 – 0.20 F_4	0.9643	0.366	0.321	108	0.9444
Aromatics	40	logP = 3.86 + 0.62 F_1 – 0.34 F_2 – 0.25 F_7 + 0.21 F_4 + 0.21 F_8 – 0.12 F_5 – 0.10 F_6 + 0.07 F_3	0.9774	0.141	0.124	167	0.9502
Halogenated and PCBs	92	logP = 3.31 + 1.58 F_1 – 0.40 F_2 + 0.18 F_4 + 0.15 F_{10} + 0.14 F_3 + 0.14 F_5 – 0.14 F_8	0.9691	0.310	0.297	376	0.9600
Total	379	logP = 2.41 + 1.06 F_1 + 0.63 F_3 + 0.37 F_{10} – 0.28 F_6 + 0.22 F_5 – 0.21 F_4 – 0.20 F_8	0.8030	0.683	0.676	216	0.7899

Except the aldehydes and aromatics subsets, the factors selected by the correlation ranking procedures are different from those of eigenvalue ranking. The models obtained almost for all subsets resulted in high statistical qualities. These are measured by the squares of correlation coefficient (R^2 >0.96) and root mean square error (*RMS* <0.321). The higher statistical qualities obtained for the alkanes, alkenes, alkynes, ethers and aldehydes, can be attributed to the lower number of compounds in these subsets. It should be noted that the results obtained by the CR–PCR procedure were better than EV–PCR. Therefore, the results of the latter are not included in Table 3.

The usefulness of QSPR models is not just their ability to reproduce known data, but also their ability to have a good estimation for any internal sample [33]. The predictive abilities of models are strongly affected by the over–fitting problem. In QSPR analyses, over–fitting problem is obtained when uninformative variables enter to the models. Another source of over–fitting is the use of exceeded number of factors in PCA–based regression methods such as PCR and PLS. There are several methods in use to estimate the quality of the models [34–36]. Cross validation is the most frequently used validation methods. Therefore, to further check the prediction ability and overfitting of the resulting models, the leave–one–out cross validation (LOO–CV) procedure was applied. In LOO–CV procedure, *n–1* sample from a total data set of each subset were used to construct a calibration set (assessment set) and to build a QSPR model between the PCs and the examined logP, and the logP property of the left out sample was estimated by the designed model. This procedure was repeated until every sample in the total data set for each subset was used for a prediction. Then, PRESS (the predicted residual sum of squares) and SSD (the sum of the squared deviation from the mean) were calculated for each regression equation. The squared correlation coefficient for cross validation (R^2_{CV}) was then calculated by the following equation $R^2_{CV} = 1 - (PRESS/SSD)$. The results of LOO–CV examination for each subset of organic compounds are listed in column 8 in Table 3. The cross–validation results show that all models (regression expressions) presented in the Table 3 have R^2_{CV} values greater than 0.92 excepted for the subset of aldehydes that it is due to small number of molecules in this class; therefore, all are reasonable QSPR models. Thus, the cross–validation test indicates that the Sh indices can model the logP of all classes of organic compounds were used in this studies, perfectly.

In the last row of Table 3 the CR–PCR model obtained for the logP of entire set of compounds by the correlation ranking procedure is listed. The trend of the PCs in order of decreasing their correlation is PC1 > PC3 > PC10 > PC6 > PC5 > PC4 > PC8 which was not in the same direction as their decreasing eigenvalue. The resulting correlation equation had correlation coefficient R^2 = 0.8030, *RMS* = 0.683, F = 216, R^2_{CV} = 0.7899. The seven factors used in this equation can explain 80.30 % of the variance in the logP of all data set of organic compounds. Further attempts were made to examine the quality of the resulted model by splitting the data set into the calibration set (322 molecules) and prediction set (57 molecules). The resulted CR–PCR model was the same as that obtained for entire set of molecules. The R^2 value and *RMS* error for the validation set are

0.7759, and 0.689, respectively. This means that the seven PCs selected by correlation ranking procedure can explain at least 77.59% variance in logP values of the external data. The results obtained with this method are presented in Table 1. The corresponding graph of calculated vs experimental logP is given in Figure 2 and the statistical parameters for the best–fitted model are represented in Table 4.

To show the superiority of modeling by PCR over the MLR, we also developed a multilinear QSPR equation for logP data by using the original Sh indices. The procedure was similar to that used in the PCR methods except that the input variables are the Sh indices and not their extracted factors. The stepwise selection of variables produced the following seven–parameter equation for the entire set of compounds:

$$\begin{gathered}\log P = -2.06 + 4.69\,Sh3 + 4.45\,Sh6 - 11.21\,Sh8 + 7.80\,Sh9 - 0.39\,Sh5 - 3.60\,Sh4 - 1.22\,Sh2 \\ N = 379,\ R^2 = 0.769,\ R^2_{CV} = 0.735,\ Se = 0.752 \text{ and } F = 176\end{gathered} \quad (15)$$

The statistical quantities of this model are also listed in Table 4. It can be seen that the number of Sh indices used by this equation is equal to the number of PCs used in the PCR method. Comparison the statistical quantity of this MLR model with those of PCR method (last row of Table 3) reveals that improved model was obtained by orthogonalization of the original Sh indices. This may be due to the some collinearity between the Sh indices.

3.2 Nonlinear Modeling

Once valid linear models were not found using PCR, steps were taken to see if prediction results could be improved by the use of artificial neural networks (ANNs). Typically, superior models can be found using ANN because they implement nonlinear relationships and because they have more adjustable parameters than the linear models. Therefore, in this study we suggested the use of ANN as the nonlinear model. A fully connected, three–layered feed–forward ANN model with back–propagation [37–39] learning algorithm is developed for nonlinear modeling between the selected PCs by the CR–PCR model. The seven PCs were test with several ANN architectures, the ANN model was confined to a single hidden layer and a sigmoid transfer function, as a more versatile transfer function, was used in this layer. Linear transfer function was used in the layer.

Because of the large number of adjustable parameters, it is possible to over–train the network. If over–training does occur, contributions of a small subset of the training set compounds may be considered as a major contribution, thus hindering the ability of the network to accurately predict the physical property in question. To avoid over–training, the data set is split into a calibration set, a prediction set and a validation set. Each connection in the network is made up of a weighting factor and a bias term. The weights and biases are changed during training based on the RMS error of the prediction set; the corresponding values are then calculated for the validation set for each of configuration. The convergence criterion was the least RMS error in the prediction set. The number

of iterations for convergence was between 15000 and 20000. In each ANN, the neuron architecture (*i.e.*, the number of nodes in hidden layer; n_H) and parameters (*i.e.*, learning rate and momentum) were optimized to reach the lowest the *RMS* error of the validation set as the performances of the resulted models, because it is believed that overtraining occurs when the *RMS* error begins to rise. At this point, the values of the weights and biases are not changed further. A plot of *RMS* error as a function of linear rate and momentum in three different numbers of nodes in hidden layer is shown in Figure 3. The results indicate that an ANN with seven PCs as input variables, 5 nodes in its hidden layer, learning rate of 0.45, and momentum of 0.75 resulted in the optimum network model. The predicted values of logP resulted from application correlation ranking ANN procedures model (CR–ANN) are shown in Table 1 and are plotted in Figure 4 against the corresponding experimental values, and the statistical parameters for the best–fitted model are represented in Table 4. As it is observed, the models obtained by the PC–ANN have superior qualities relative to those obtained by PCR. This means that there are nonlinear relationships between the proposed Sh topological indices and the lipophilicity of the organic molecules used in this study. A comparison between the results obtained by the eigenvalue ranking and correlation ranking–based PC–ANN models revealed that the latter produced more accurate results, which is in accordance with our previous findings [16,30,31].

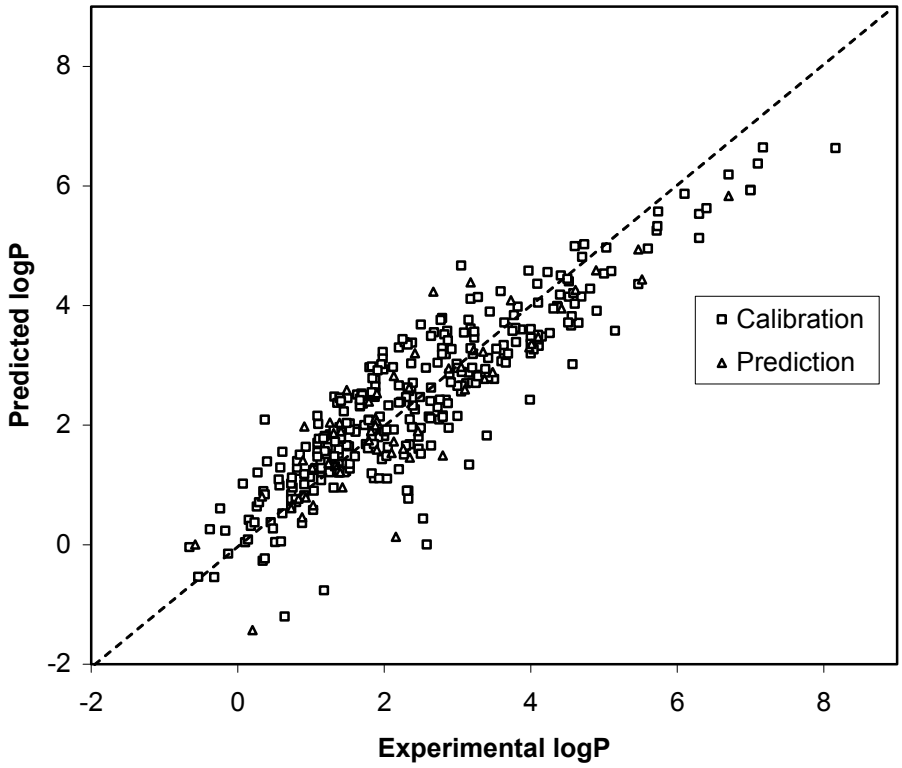

Figure 2. Plot of the predicted logP by CR–PCR against the experimental values. The dash line is the ideal fit to the straight line.

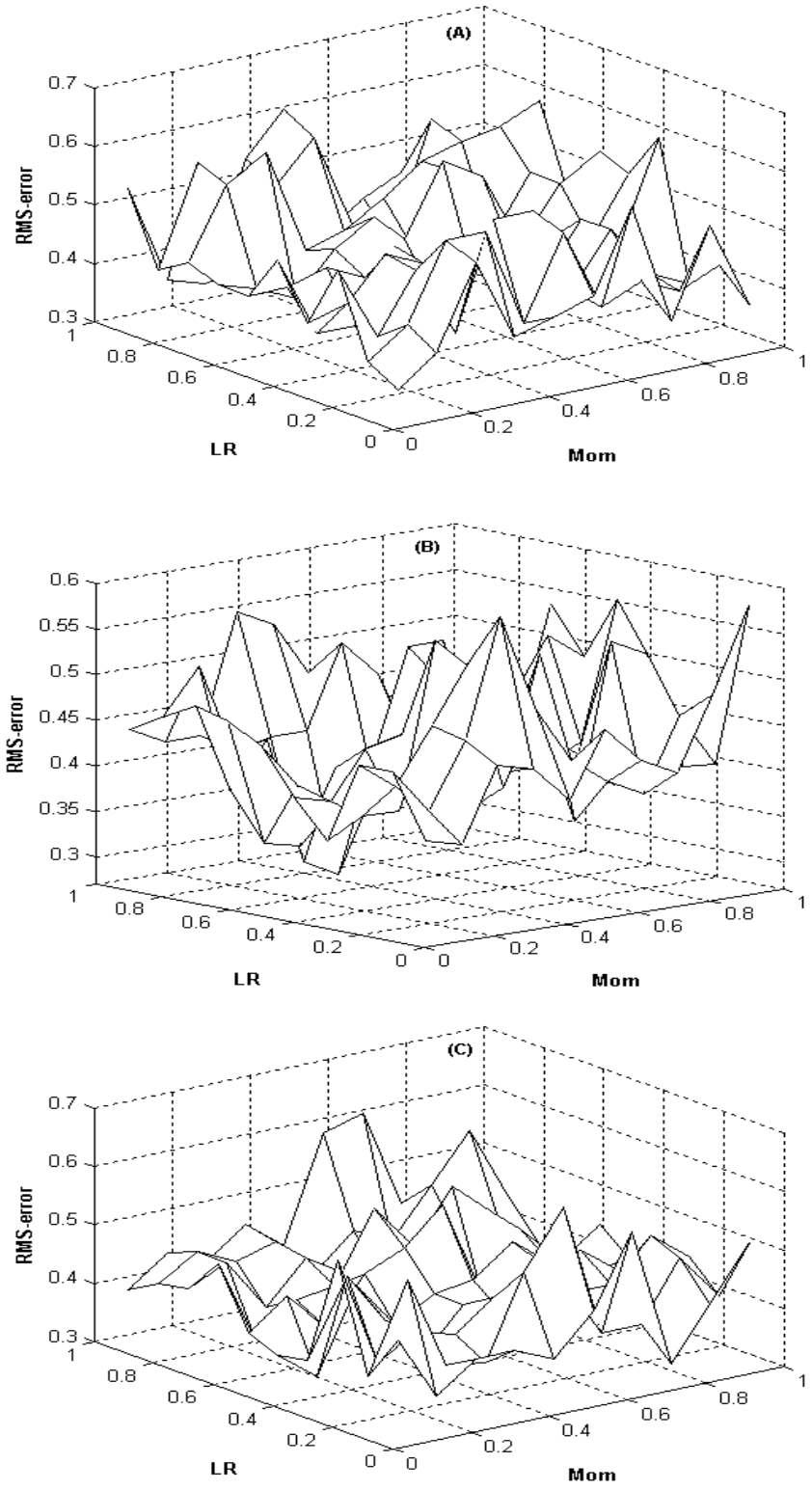

Figure 3. Optimization of linear rate (LR), momentum (Mom) and number of hidden layer nodes (n_H) for ANN modeling; (**A**) $n_H = 4$; (**B**) $n_H = 5$ and (**C**) $n_H = 6$.

Modeling of nonlinear relationship between the original Sh indices and logP was also performed to check the superiority of the orthogonalization of the original variable. The input of the ANN was those Sh indices that selected by MLR analysis (equation 15). The results are summarized in Table 4. The data clearly investigate better quality of the results obtained by PC–ANN.

Table 4. Statistical parameters of the models used for calculating partition coefficients

Parameter	Linear modeling			
	MLR		PCR	
	Calibration	Prediction	Calibration	Prediction
N	322	57	322	57
SE	0.755	0.783	0.677	0.689
RMS	0.732	0.754	0.675	0.689
R^2	0.7690	0.7431	0.8076	0.7758
F	176	152	188	190
Error range	(–1.74)–(2.44)	(–1.88)–(2.53)	(–1.72)–(2.58)	(–1.56)–(2.03)

Parameter	Non–linear modeling					
	MLR–ANN			PC–ANN		
	Calibration	Prediction	Validation	Calibration	Prediction	Validation
N	275	57	47	275	57	47
SE	0.345	0.361	0.427	0.224	0.281	0.271
RMS	0.355	0.362	0.432	0.224	0.281	0.272
R^2	0.9489	0.9377	0.8995	0.9798	0.9626	0.9605
F	5065	829	401	13240	1415	1094
Error range	(–1.26)–(1.42)	(0.90)–(1.07)	(–0.79)–(1.61)	(–0.74)–(0.77)	(–0.74)–(0.65)	(–0.73)–(0.61)

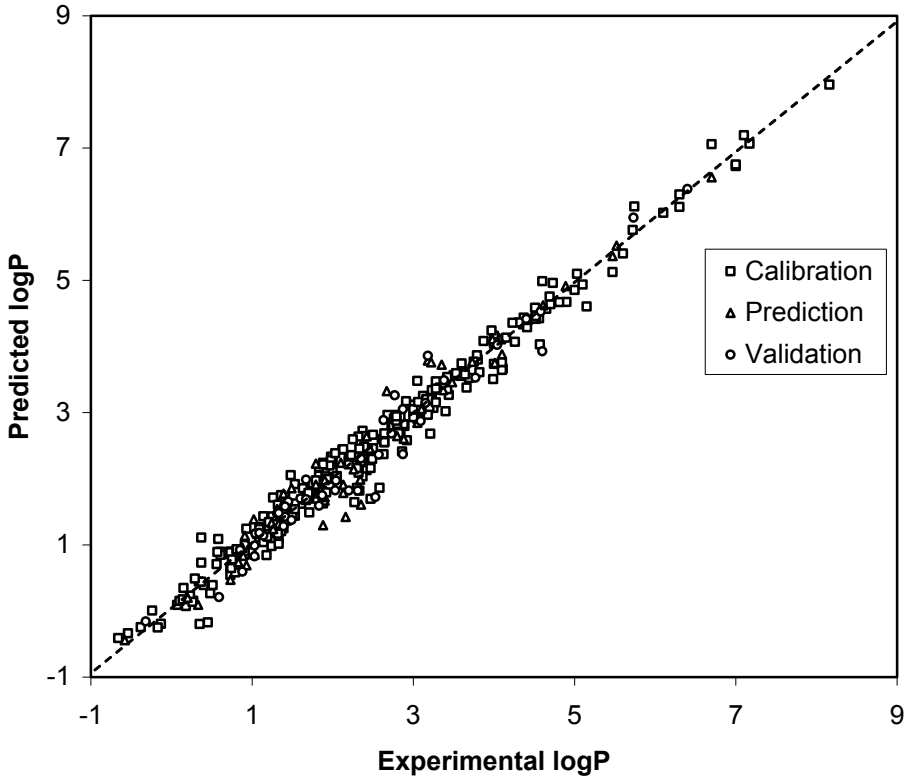

Figure 4. Plot of the predicted logP by CR–ANN against the experimental values. The dash lines are the ideal fit to the straight line.

A comparison between the four types of the modeling methods used in this study revealed that the relationship between the proposed Sh topological indices and logP of chemicals is modeled the best by the nonlinear method. Both the MLR–ANN and PC–ANN produced better results relative to MLR and PCR.

Comparison between the results obtained by the original variables and the extracted features (PCs) indicates that orthogonalization of variables enhanced the quality of QSPR modeling. This may be due to the fact that modeling with the original variables uses only the information part of the variables used in the model, while the PCs used in the modeling have some information from all of the original variables.

As it was noticed in the introduction section, different QSPR methods have been represented for predicting logP. However, some papers are concerned with a small set of compounds [41,42] or use complicated calculation procedures [43,44]. Toropov and Toropova [41] reported a modeling method for a set of 38 organic molecules by using correlation weight of connectivity index. The correlation coefficient and standard error of this model for the prediction set were 0.9649 and 0.42, respectively. The correlation coefficient of Tehrany *et al*. model [42] that uses constitutional and quantum chemical descriptors for 42 organic compounds was 0.983. It seems that the statistical quality of these models is higher than that of our model; however, ours models have been used for a large set of organics. The QSPR model proposed by Bodor and Buchwald [43] for the data set relatively similar to our data set have $R = 0.989$ for 320 molecules. However, they needed accurate three–dimensional geometry of molecules to calculate molecular volume. In addition, the predictive model developed by Eisfeld and Maurer [44] employed ab initio quantum chemical calculations that are time consuming. Although the quality of our model is comparable with the other models, its main advantage is the simplicity of the calculation of the Sh indices even with a simple calculator.

4 CONCLUSIONS

The usefulness of the some newly proposed topological indices (Sh indices) in quantitative structure–lipophilicity relationship analysis of the n–octanol/water partition coefficient (logP) of 379 diverse organic compounds by using the principal component regression and principal component–artificial neural network modeling methods was examined. The PCs were entered to the models based on their decreasing eigenvalues (EV) and their correlation ranking coefficients (CR) with the logP, in which the latter produced better results. PCR analysis of the data showed that proposed Sh indices could explain about 80% of variations in the logP data; while the variations explained by the ANN modeling were more than 96%. These results confirm the suitability of the indices in QSPR analysis of the lipophilicity data. The advantages of the proposed method are ease of calculation of the topological indices (even with a simple calculator) and the good predictive ability.

5 REFERENCES

[1] J. Gasteiger, *Handbook of Chemoinformatics: From Data to Knowledge*, *4 Volume Set*, 2003.
[2] R. Todeschini and V. Consonnu, *Handbook of Molecular Descriptors*, Wiley–VCH, Weinheim, 2000.
[3] M. Shamsipur, B. Hemmateenejad, and M. Akhond, Highly Correlating Distance/Connectivity–Based Topological Indices. 1: QSPR Studies of Alkanes, *Bull*. *Korean Chem*. *Soc*. **2004**, *25*, 1–7.
[4] M. Shamsipur, R. Ghavami, B. Hemmateenejad, and H. Sharghi, Highly Correlating Distance/Connectivity–Based Topological Indices. 2: Prediction of 15 Properties of a Large Set of Alkanes Using a Stepwise Factor Selection–Based PCR Analysis, *QSAR Comb*. *Sci*. **2004**, *23*, 734–753.
[5] (a) C. Hansch, and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979. (b) C. Hansch, A. Leo, and D. Hoekman, *Exploring QSAR – Hydrophobic*, *Electronic*, *and Steric Constants*; ACS Prefessional Reference Book; American Chemical Society: Washington, DC, 1995.
[6] V. K. Agrawal, M. Gupta, J. Singh and P. V. Khadikar, A novel method of estimation of lipophilicity using distance–based topological indices: dominating role of equalized electronegativity, *Bioorg*. *Med*. *Chem*. **2005**, *13*, 2109–2120.
[7] D. Eros, I. Kovesdi, L. Orfi, K. Takacs–Novak and G. Kier, Reliability of logP predictions based on calculated molecular descriptors: a critical review, *Curr*. *Med*. *Chem*. **2002**, *9*, 1819–1829.
[8] G. E. Kellogg and D. J. Abraham, Hydrophobicity: is LogP (o/w) more than the sum of its parts?, *Eur*. *J*. *Med*. *Chem*. **2000**, *35*, 651–661.
[9] L. Molnar, G. M. Keseru, A. Papp, Z. Gulyas and F. Darvas, A Neural Network Based Prediction of Octanol–Water Partition Coefficients Using Atomic5 Fragmental Descriptors, *Bioorg*. *Med*. *Chem*. *Lett*. **2004**, *14*, 851–853.
[10] J. K. Wegner and A. Zell, Prediction of Aqueous Solubility and Partition Coefficient Optimized by a Genetic Algorithm Based Descriptor Selection Method, *J*. *Chem*. *Inf*. *Comput*. *Sci*. **2003**, *43*, 1077–1084.
[11] D.C. Montgomery, D. C.; Peck, E. A. *Introduction to Linear Regression Analysis*, Wiley, New York, 1982.
[12] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.
[13] J. H. Kalivas, and P. M. Lang, *Mathematical Analysis of Spectral Orthogonality*, Marcel Dekker, New York, 1994.
[14] G. Puchwein, Selection of Calibration Samples for Near–Infrared Spectrometry by Factor Analysis of Spectra, *Anal*. *Chem*. **1988**, *60*, 569–573.
[15] B. Hemmateenejad, M. Akhond, R. Miri and M. Shamsipur, Genetic Algorithm Applied to the selection of Factors in Principal Component–Artificial Neural Networks: Application to QSAR Study of Calcium Channel Antagonist Activity of 1,4–Dihydropyridines (Nifedipine Analogous), *J*. *Chem*. *Inf*. *Comput*. *Sci*. **2003**, *43*, 1328–1334.
[16] B. Hemmateenejad, Correlation Ranking Procedure for Factor Selection in PC–ANN Modeling and Application to ADMETox Evaluation, *Chemom*. *Intell*, *Lab*. *Syst*. **2005**, *75*, 231–245.
[17] B. Hemmateenejad, Optimal QSAR Analysis of the Carcinogenic Activity of Drug by Correlation Ranking and Genetic Algorithm–Based PCR, *J*. *Chemomet*. **2004**, *1*8, 475–485.
[18] M. Shamsipur, B. Hemmateenejad, and M. Akhond, Multicomponent Acid–Base Titration by Principal Component–Artificial Neural Network Calibration, *Anal*. *Chim*. *Acta* **2002**, *461*, 147–153.
[19] G. Schneider, and P. Wrede, Artificial Neural Networks for Computer–Based Molecular Design, *Prog*. *Biophys*. *Mol*. *Biol*. **1998**, *70*, 175–222.
[20] P. J. Gemperline, J. R. Long, and V. G. Gregoriou, Nonlinear Multivariate Calibration Using Principal Components Regression and Artificial Neural Networks, *Anal*. *Chem*. **1991**, *63*, 2313–2317.
[21] L. B. Kier, Indexes of Molecular Shape From Chemical Graphs, Acta *Pharm*. *Jugosl*. **1986**, *36*, 171–188.
[22] J. Sangster, *Octanol–Water Partition Coefficients: Fundamentals and Physical Chemistry*, New York: Wiley, 1997.
[23] C. Hansch, D. Kim, A. J. Leo, E. Novellino, C. Silipo, and A. Vittoria, Toward a Quantitative Toxicology of Organic Compounds, *Crit*. *Rev*. *Toxicol*. **1989**, *19*, 185–226.
[24] R. F. Rekker, and H. M. De Kort, H. M., The Hydrophobic Fragmental Constant, an Extension to a 1000 Datapoint Set, *Eur*. *J*. *Med*. *Chem*. **1979**, *14*, 479–488.
[25] E. Wolfgang, and M. Gerd, Study on the Correlation and Prediction of Octanol/Water Partition Coefficients by Quantum Chemical Calculations, *J*. *Phys*. *Chem*. *B*. **1999**, *103*, 5716–5729.
[26] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
[27] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press Ltd.: Letchworth, England, 1986
[28] E. R. Malinowsky, *Factor Analysis in Chemistry*; John Wiley: New York, 2000.
[29] I. N. Koprinarov, A. P. Hitchcock, C. T. McCrory, and R. F. Childs, Quantitative Mapping of Structured

Polymeric Systems Using Singular Value Decomposition Analysis of Soft x–ray Images, *J. Phys. Chem. B.* **2002**, *106*, 5358–5364.
[30] B. Hemmateemejade, and M. Shamsipur, Quantitative Structure–Electrochemistry Relationship Study of Some Organic Compounds Using PC–ANN and PCR, *Int. Elect. J. Mol. Des.* **2004**, *3*, 316–334.
[31] B. Hemmateemejade, M. A. Safarpour, R. Miri, N. Nesari, Toward an Optimal Procedure for PC–ANN Model Building: Prediction of the Carcinogenic Activity of a Large Set of Drugs, *J. Chem. Inf. Model*. in press.
[32] B. Hemmateemejade, M. A. Safarpour, R. Miri and F. Taghavi, Application of *ab initio* Theory to QSAR Study of 1,4–Dihydropyridines–Based Calcium Channel Blockers Using GA–MLR and PC–GA–ANN Procedures, *J. Comput. Chem.* **2004**, *25*, 1495–1503.
[33] P. Gramatica, and E. Papa, QSAR Modeling of Bioconcentration Factor by Theoretical Molecular Descriptors, *QSAR Comb. Sci.* **2003**, *22*, 374–385.
[34] R. D. Clark, D. G. Sprous, and J. M. Leonard, *Rational Approaches to Drug Design*, Edi; Höltje, H.–D. and Sippl, W., pp. 475–485. Prous Science, Barcelona.
[35] S. Wold, and L. Eriksson, *Chemometric Methods in Molecular Design– Vol 2*, *Ed.*; *van de Waterbeemd*, *H.*, pp. 309–318. VCH Publishers, New York (NY), **1995**.
[36] S. Wold, *Quantitative Structure–Activity Relationships*, **1991**, *10*, 191–193.
[37] D. E. Rumelhart, G. E. Hinton, and. R. J. Williams, Learning Representation by Back–Propagating Errors, *Nature*, **1986**, *323*, 533–539.
[38] D. Svozil, V. Kvasnicka and J. Pospichal, Introduction to Multi–layer Feed–forward Neural Networks, *Chemomet. Intell. Lab. Syst.* **1997**, *39*, 43–62.
[39] J. Devillers, D. Domie and C. Guillon, Autocorrelation Modeling of Lipophilicity with a Back–propagation Neural–Network, *Eur. J. Med. Chem.* **1998**, *39*, 659–664.
[40] J. Taskinen and J. Yliruusi, Prediction of Physicochemical Properties Based on Neural Network Modeling, *J. Adv. Deliv. Rev.* **2003**, *55*, 1163–1183.
[41] A. A. Toropov, A. P. Toropopva, Modeling of lipophilicity by means of correlation weighting of local graph invariants, *J. Mol. Struct.* **2001**, 538, 197–199.
[42] E. A. Tehrany, F. Fournier, S. Desobry, Simple method to calculate octanol–water partition coefficient of organic compounds, *J. Food Eng.* **2004**, 64, 315–320.
[43] N. Bodor, P. Buchwald, Molecular size based approach to estimate partition properties for organic solutes, *J. Phys. Chem. B* **1997**, 101, 3404–3412.
[44] W. Eisfeld, G. Maurer, Study on the correlation and prediction of octanol/water partition coefficient by quantum chemical calculations, *J. Phys. Chem. B* **1999**, 103, 5716–5729.