ELSEVIER

Journal of Chromatography A, 955 (2002) 273–280

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Simultaneous modeling of the Kovats retention indices on OV-1 and SE-54 stationary phases using artificial neural networks

M.H. Fatemi

Department of Chemistry, Mazandaran University, P.O. Box 453, Babolsar, Iran

Received 21 November 2001; received in revised form 14 February 2002; accepted 14 February 2002

Abstract

In this study, a quantitative structure–property relationship technique has been used for the simultaneous prediction of Kovats retention indices for some esters, alcohols, aldehyde and ketones on OV-1 and SE-54 stationary phases, using an artificial neural network (ANN). The best-selected descriptors that appear in the models are the molecular values, number of atoms in each molecule, molecular shadow area on the *xy* plane and the energy level of the highest occupied molecular orbital. A 4-6-2 ANN was generated using these descriptors as inputs and its outputs will be the Kovats retention indices on OV-1 and SE-54 stationary phases. After optimization of the network parameters, the network was trained using a training set. For the evaluation of the predictive power of the generated ANN, an optimized network was used to predict the Kovats retention indices of the prediction set. The results obtained in this study showed that the average percentage deviation between the predicted ANN and the experimental values of Kovats retention indices for the prediction set were 2.5 and 3.0% on the OV-1 and SE-54 stationary phases, respectively. These values are in good agreement with the experimental results.

Keywords: Retention indices; Regression analysis; Mathematical modelling; Neural networks, artificial; Molecular descriptors; Quantitative structure–property relationships; Alcohols; Esters; Carbonyl compounds

1. Introduction

The Kovats retention indices in gas chromatography (GC) represent the retention behavior of a compound relative to a standard set of hydrocarbons, using a logarithmic scale [1]. The identification of many compounds is often accomplished on the basis of GC peak comparison with a standard sample of the suspected material. However, it is not always possible to obtain samples of pure standard compound for such comparison, therefore the development of a theoretical model for estimating the retention indices seems to be necessary. Retention is a phenomenon that primarily depends on the interactions between the solute molecules and the stationary phase. The forces associated with these interactions can be related to the geometric and topological structure and also to the electronic environments of a molecule. Quantitative structure–property relationships (QSPRs) have been demonstrated to be a powerful tool in chromatographic studies [2]. QSPRs have been used to obtain simple models to explain and predict the chromatographic behavior of various classes of compounds. Mekenyan et al. derived linear quantitative retention–structure models in gas chromatography for 41 alkylbenzenes using boiling point,

E-mail address: mhfatemi@umz.ac.ir (M.H. Fatemi).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00169-3

two geometric indices and two electronic structure parameters as descriptors [3]. Woloszyn and Jurs correlated the observed Kovats retention indices of sulfur vesicants by multiple linear regression (MLR) techniques using nine topological, electronic and geometrical descriptors [4]. They also predicted the GC retention behavior of 67 hydrocarbons [5]. Collantes et al. studied the chromatographic retention data of 31 unsubstituted 3-6 ring polycyclic aromatic hydrocarbons using CoMFA [6]. Katritzky et al. also reported the QSPR modeling of the GC retention indices of methyl-branched hydrocarbons produced by insects [7]. Some other reports in this area are listed in Refs. [8–11].

Artificial neural networks (ANNs) have been used for a wide variety of chemical problems such as prediction of ^{13}C NMR chemical shift [12] and selectivity coefficients of ion-selective electrodes [13], simulation of mass spectra [14] and modeling of ion-interaction chromatography [15,16]. ANNs were also used in quantitative structure–activity relationship studies [17–21].

In this report, an ANN was employed to generate a QSPR model between the molecular-based structural parameters and observed retention indices of some oxygenated organic compounds on OV-1 and SE-54 stationary phases simultaneously.

2. Experimental

2.1. Data set

The data set of Kovats retention indices was taken from the values reported by Zhang et al. [22]. The molecules in the data set including esters, ketones, aldehydes and alcohols, are shown in Table 1. The Kovats retention indices of all molecules included in the data set were obtained under the same conditions on two stationary phases; OV-1 (dimethylpolysiloxane) and SE-54 (5% phenyl–95% dimethylpolysiloxane). The Kovats retention indices fall in the range of 605.79–986.60 for propyl formate and 3,6-dimethyl-3-heptanone on OV-1 stationary phase, respectively, and in the range of 623.60 for propyl formate to 1000 for 3,6-dimethyl-3-heptanone and ethyl hexanate on SE-54 stationary phase.

The data set was randomly divided into two groups; a training set and a prediction set consisting of 87 and 20 molecules, respectively. The training set was used for the model generation and the prediction set was used for the evaluation of the generated model.

2.2. Descriptors generation

Retention in gas chromatography is the result of competitive solubility of the solute between the mobile and stationary phases. The molecular structure and chemical properties of the solute determine the type and extent of the interactions of the solute with these phases. The differences between these properties govern the retention behavior through the column. Due to the diversity of the molecules studied in this work, 64 different descriptors were calculated. These parameters encoded different aspects of the molecular structure and consist of electronic, geometric and topological descriptors. Geometric descriptors were calculated using optimized Cartesian coordinates and the van der Waals radius of each atom in the molecule [23,24]. Electronic descriptors were calculated using the MOPAC program (version 6) [25]. Topological descriptors were calculated using two-dimensional representation of the molecules. Some of the descriptors generated for each compound encoded similar information about the molecule of interest. Therefore, it was desirable to test each descriptor and eliminate those that show high correlation ($R>0.95$) with each other. A total of 23 out of 64 descriptors showed high correlation and were removed from the next consideration. Subsequently, the method of stepwise multiple linear regression was used for selection of important descriptors. The descriptors that appear in the best MLR equations for OV-1 and SE-54 stationary phases are identical and are shown in Table 2. These descriptors were used as inputs for the generated ANN.

2.3. Artificial neural network generation

A detailed description of the theory behind a neural network has been adequately described elsewhere [26–28]. In addition we reported some relevant principles of the ANNs in previous papers [12,19–21]. The program for the feed-forward neural

Table 1
Data set and corresponding observed (obs) and ANN predicted (cal) values of the retention indices (I)

No.		$(I_{ov\text{-}1})_{obs}$	$(I_{ov\text{-}1})_{cal}$	$(I_{SE\text{-}54})_{obs}$	$(I_{SE\text{-}54})_{cal}$
Training set					
1	3,3-Dimethyl-1-butanol	778.77	750.73	763.63	761.69
2	3-Methyl-3-hexanol	826.62	806.45	841.11	821.67
3	2,2,4-Trimethyl-3-pentanol	881.49	909.19	894.00	926.65
4	4-Methyl-1-pentanol	821.19	825.56	836.97	845.20
5	2-Pentanone	666.34	667.14	687.79	685.86
6	Isopropyl acetate	640.54	644.68	661.78	662.13
7	Propyl formate	605.79	623.61	623.60	634.78
8	Isobutyl formate	673.40	683.85	689.84	701.58
9	4-Ethyl-3-hexanol	953.26	960.86	967.63	970.93
10	Butyl formate	707.64	698.36	725.53	718.35
11	2,4-Dimethyl-2-pentanol	775.91	810.96	789.03	826.48
12	2-Hexanone	761.93	863.63	790.03	882.85
13	1-Heptanol	955.05	911.87	971.73	926.58
14	2-Methyl-3-Pentanol	757.96	783.44	772.28	797.27
15	Isobutyl propionate	852.83	888.02	869.02	905.18
16	2-Ethyl hexanal	934.65	868.92	954.71	888.53
17	2,2,4,4-Tetramethyl-3-pentanone	900.00	911.12	914.09	928.84
18	3-Methyl-3-butene-1-ol	713.62	712.77	731.70	739.13
19	3,3-Dimethyl-2-butanone	693.05	718.32	711.58	731.17
20	Butyl butyrate	979.36	981.77	997.07	987.35
21	2-Methyl-3-hexanone	819.95	825.91	838.42	844.22
22	2,2-Dimethyl-1-propanol	657.34	642.70	670.46	652.47
23	2-Methyl-2-heptanol	916.43	938.26	930.38	952.08
24	Methyl propionate	615.21	623.11	630.43	633.96
25	2-Amino-1-butanol	804.30	779.83	830.50	799.17
26	Methyl isobutyrate	670.97	658.76	686.58	671.75
27	3,6-Dimethyl-3-heptanol	986.60	959.14	1000.00	969.73
28	3-Methyl-2-butanone	640.90	641.72	661.44	654.14
29	3-Methyl-1-butanol	719.03	734.20	734.39	753.32
30	2,2-Dimethyl-3-pentanol	805.63	829.81	818.97	846.42
31	2-Ethylbutyl acetate	956.99	960.48	974.66	970.47
32	Isobutyl isobutyrate	900.00	898.15	915.56	915.79
33	Methyl hexanoate	907.01	893.52	925.46	911.06
34	2-Methyl-4-pentene-2-ol	694.62	722.19	710.28	735.91
35	6-Methyl-2-heptanol	951.10	968.57	965.00	977.02
36	3-Heptanone	865.79	864.66	886.89	884.16
37	2-Methyl pentanal	742.38	757.47	762.95	778.87
38	3-Pentanone	676.41	675.80	700.00	695.74
39	Butyl isobutyrate	938.55	965.79	954.26	974.81
40	Ethyl hexanate	982.90	967.49	1000.00	976.24
41	2-Methyl-2-pentanol	717.57	768.21	731.39	780.57
42	Pentyl acetate	896.36	891.12	914.88	908.30
43	2-Ethyl-1-butanol	825.94	781.74	841.00	795.55
44	Propyl butyrate	881.53	890.74	898.88	907.96
45	2-Octanone	968.77	956.29	991.27	967.11
46	2-Propanol	700.00	701.67	725.82	725.48
47	2,4-Dimethyl-3-pentanone	779.01	774.04	795.28	787.94
48	4-Methyl-1-pentene 3-ol	740.10	739.78	754.41	758.00
49	Ethyl isovalerate	838.35	855.63	854.28	873.55
50	Butyl acetate	796.18	808.25	814.16	830.20
51	Methyl butyrate	705.61	707.93	722.96	729.25
52	3-Butene-1-ol	618.91	669.16	636.98	697.29
53	2-Methyl butanal	636.32	645.06	657.70	658.18
54	5-Methyl-3-heptanol	943.58	940.62	957.88	954.12

Table 1. Continued

No.		$(I_{ov\text{-}1})_{obs}$	$(I_{ov\text{-}1})_{cal}$	$(I_{SE\text{-}54})_{obs}$	$(I_{SE\text{-}54})_{cal}$
55	2-Pentanol	682.66	675.90	700.00	695.87
56	4-Heptanol	875.42	864.68	890.00	884.16
57	2-Amino-2-methyl-1-propanol	730.84	745.68	754.91	763.77
58	2-Butene-1-ol	649.00	666.17	666.00	697.23
59	3-Methyl-2-butanol	666.02	668.77	680.26	681.32
60	2,3-Dimethyl-4-pentene-2-ol	781.91	774.41	797.12	787.59
61	3-Methyl-2-pentanone	734.76	745.40	754.92	764.01
62	4-Heptanone	853.35	864.68	873.44	884.16
63	2-Methyl-2-hexanol	817.33	843.95	831.38	860.48
64	2-Methyl-2-butanol	626.20	660.73	640.33	672.48
65	Ethyl butyrate	784.04	810.90	800.00	832.58
66	2-Ethyl-4-methyl-1-pentanol	972.00	920.30	986.00	936.48
67	3-Hexanone	764.84	760.51	785.88	782.70
68	2,2,4-Trimethyl-1-pentanol	930.00	928.32	943.00	943.49
69	Isobutyl acetate	757.65	763.23	774.13	789.12
70	1-Hexanol	852.96	833.12	869.44	850.01
71	3-Ethyl-3-pentanol	843.09	841.96	858.19	858.95
72	2,2-Dimethyl-1-heptanol	867.57	936.30	881.00	950.81
73	1-Heptene-4-ol	850.88	869.22	867.52	889.40
74	4-Methyl-2-pentanol	744.14	746.33	758.42	765.08
75	Methyl isovalerate	761.30	745.53	777.34	758.71
76	5-Methyl-3-hexanol	838.15	815.86	852.08	832.07
77	3-Amino-1-propanol	775.50	805.15	807.76	827.64
78	2,2-Dimethyl-3-heptanone	964.66	924.43	980.56	940.63
79	2,3-Dimethyl-3-pentanol	823.66	810.58	838.69	826.17
80	4-Heptene-1-ol	735.81	668.51	754.51	696.40
81	Isobutyl alcohol	611.31	621.81	626.00	638.88
82	2-Methyl-1-pentanol	818.35	821.81	833.58	838.02
83	2,4-Dimethyl-3-heptanol	821.18	818.54	834.49	834.73
84	1-Butanol	646.48	628.91	662.08	640.80
85	5-Methyl-2-hexanone	836.53	862.32	858.37	881.47
86	5-Methyl-3-hexanone	816.74	813.38	835.88	830.86
87	2-Heptanol	885.57	904.59	900.00	919.86
Prediction set					
1	Propyl acetate	696.34	688.89	713.63	704.44
2	Ethyl propionate	694.19	701.50	711.16	720.20
3	Butyl propionate	891.40	887.80	909.12	904.20
4	4-Methyl-2-pentanone	721.24	743.43	741.61	758.95
5	2-Heptanone	868.70	858.26	891.01	876.13
6	3-Pentanol	684.21	698.77	700.00	713.40
7	3-Methyl-1-pentanol	828.82	777.77	845.00	790.51
8	4-Octanol	975.50	955.34	990.22	998.13
9	2-Methyl-2-propen-1-ol	629.11	608.20	646.35	672.16
10	2-Methyl-1-pentene-3-ol	763.67	774.13	778.54	797.54
11	2-Amino ethanol	644.27	651.05	670.38	670.08
12	Propyl propionate	792.58	765.15	809.79	782.02
13	1-Pentanol	750.40	718.38	766.59	735.49
14	Isobutyl butyrate	940.26	962.03	956.57	971.66
15	2-Methyl-3-pentanone	733.02	735.18	752.40	749.50
16	2,6-Dimethyl-4-heptanone	954.66	932.61	970.95	947.69
17	2,3-Dimethyl-2-butanol	715.26	667.94	729.44	684.78
18	3-Hexanol	780.36	798.56	795.07	812.80
19	3,5-Dimethyl-3-hexanol	883.13	850.25	896.48	932.84
20	3-Octanol	981.97	998.12	997.47	968.51

Table 2
Specification of multiple linear regression models

Descriptor	Notation	Coefficient of OV-1	Coefficient of SE-54
1. Number of atoms in molecule	NA	25.202 (±2.237)	25.154 (±2.269)
2. Molecular volume	MV	9.548 (±1.315)	10.150 (±1.334)
3. Molecular shadow area on *xy* plane	S_{xy}	−2.695 (±0.536)	−2.876 (±0.544)
4. Highest occupied molecular orbital	HOMO	41.323 (±11.243)	45.498 (±11.406)
Constant		659.740 (±114.520)	720.472 (±116.183)

network that was trained by back-propagation algorithm was written in FORTRAN 77. This network has four nodes in the input layer and the numbers of nodes in the output layer were set to be two. The signals of each node in the output layer represent the Kovats retention indices on OV-1 and SE-54 stationary phase separately. The number of nodes in the hidden layer would be optimized. The initial weights were randomly selected from a uniform distribution that ranged between −0.3 and +0.3. The initial bias values were set to be one. These values were optimized during the network training. The value of each input was divided into its mean value to bring the values of the input variables into the dynamic range of the sigmoid transfer function in the ANN. Before training, the network was optimized for the number of nodes in the hidden layer, learning rates and momentum. Then the network was trained using the training set by back-propagation strategy to optimize the values of the weights and biases. It is known that neural networks can become over-trained. An over-trained network has usually learned perfectly the stimulus pattern it has seen (training set) but cannot give accurate predictions for unseen stimuli, and it is no longer able to generalize. There are several methods for overcoming this problem. One of the superior methods is to use a test set to validate the prediction power of the network during its training [21]. In order to evaluate the performance of the ANN, standard error of calibration (SEC) and standard error of prediction (SEP) were used [29].

3. Results and discussion

The data set and corresponding observed and ANN predicted values of the Kovats retention indices of all molecules studied in this work are shown in Table 1. The molecules included in the data set are alcohols, esters, aldehydes and ketones. All of these types of molecules are included in the prediction set. Therefore, although the molecules included in the prediction set are chosen randomly, they adequately represent the training set. Table 2 shows the best MLR models. It can be seen from this table that four identical descriptors are used in the two MLR models. These variables encode different aspects of the molecular structure. Among the different factors affecting the retention behavior of molecules, mass, size and bulkiness are most important. The number of atoms in each molecule, the molecular volume and the molecular shadow area on the *xy* plane represent these properties that can appear in models. The appearance of the energy level of the highest occupied molecular orbital in the models represents the role of electronic interaction between a solute molecule and stationary phase. The calculated values of these descriptors appearing in the best models are shown in Table 3 for whole molecules included in the training and prediction sets.

The next step was the generation of the artificial neural network. Before training the network, the parameters of the number of nodes in the hidden layer, weights and biases learning rates and momentum values were optimized. The procedure for the optimization of these parameters was reported in our previous papers [14,21]. Table 4 shows the architecture and specifications of the optimized ANN. After the optimization of the ANN's parameters the network was trained using a training set for the adjustment of weights and bias values. To control the overfitting of the network during the training procedure, the values of SEC and SEP were calculated and recorded to monitor the extent of the learning after 500 iterations. Results obtained showed that after 47 000 iterations, the value of SEP started to increase and overfitting began. To maintain the predictive power of the network at a desirable level,

Table 3
The values of the descriptors that were used in this work

No.[a]	NA	MV	S_{xy}	HOMO
Training set				
1	21	120.57	33.85	−10.76
2	24	137.42	33.84	−10.76
3	27	153.29	42.64	−10.58
4	21	121.02	40.55	−10.85
5	16	98.26	34.74	−10.53
6	17	123.18	12.87	−11.20
7	14	90.16	33.36	−11.22
8	17	105.92	36.23	−10.58
9	27	153.73	48.70	−10.63
10	17	107.05	37.84	−11.17
11	24	137.22	37.32	−10.76
12	22	131.99	44.97	−10.51
13	24	137.90	45.96	−10.85
14	21	120.80	37.92	−10.76
15	23	140.28	48.03	−11.15
16	22	132.03	45.45	−10.43
17	28	164.02	41.23	−10.10
18	16	139.37	45.69	−9.78
19	19	114.71	34.51	−10.35
20	29	174.21	58.79	−11.15
21	22	131.84	42.28	−10.31
22	18	103.93	30.33	−10.87
23	27	154.44	46.87	−10.80
24	14	89.79	33.01	−11.23
25	17	97.94	32.64	−9.75
26	17	106.72	35.17	−11.15
27	30	170.97	51.68	−10.74
28	16	98.21	31.66	−10.43
29	18	104.12	35.72	−10.81
30	24	136.91	39.61	−10.62
31	26	156.94	51.83	−11.03
32	26	157.16	46.13	−11.07
33	23	140.35	48.44	−10.95
34	19	114.97	35.16	−10.11
35	27	154.51	50.50	−10.87
36	22	131.96	44.94	−10.40
37	19	115.17	39.48	−10.44
38	16	98.20	34.64	−10.42
39	26	157.32	53.52	−11.07
40	26	157.26	53.55	−10.95
41	21	120.74	36.66	−10.84
42	23	140.47	48.54	−11.17
43	21	120.63	37.71	−10.67
44	23	140.52	48.50	−11.16
45	25	148.82	50.07	−10.51
46	12	70.51	24.68	−10.91
47	22	131.56	37.24	−10.22
48	19	114.94	37.86	−10.23
49	23	140.21	45.21	−10.81
50	20	123.74	43.39	−11.17
51	17	106.62	38.17	−11.24
52	13	81.48	31.94	−10.02
53	16	98.40	32.05	−10.42
54	27	154.05	46.85	−10.72
55	16	98.21	34.65	−10.42
56	22	131.92	44.92	−10.40
57	17	97.90	30.50	−9.78
58	13	81.51	31.95	−9.61
59	18	104.07	32.26	−10.84
60	22	131.43	32.49	−10.09
61	19	114.90	38.21	−10.42
62	22	131.92	44.92	−10.40
63	24	137.57	41.80	−10.82
64	18	103.91	31.59	−10.81
65	20	123.59	43.36	−11.21
66	27	154.19	44.96	−10.75
67	19	115.10	39.78	−10.41
68	27	153.67	45.73	−10.78
69	20	131.89	46.49	−11.38
70	21	121.06	40.81	−10.84
71	24	136.75	40.54	−10.57
72	30	170.96	49.21	−10.81
73	22	132.04	45.17	−10.19
74	19	115.11	38.38	−10.48
75	20	123.59	38.17	−11.09
76	21	120.54	39.68	−10.68
77	17	97.89	34.97	−9.67
78	28	165.23	46.43	−10.23
79	24	136.88	35.13	−10.68
80	13	81.56	31.91	−10.02
81	15	87.23	29.62	−10.86
82	21	120.84	40.13	−10.87
83	24	137.20	37.63	−10.62
84	15	87.33	30.62	−10.84
85	22	131.92	44.82	−10.51
86	22	131.96	41.69	−10.37
87	24	137.79	45.26	−10.79
Prediction set				
1	17	106.83	38.35	−11.19
2	17	106.62	38.16	−11.18
3	23	140.48	48.52	−11.15
4	19	115.11	38.05	−10.49
5	22	131.99	44.97	−10.51
6	18	104.10	34.45	−10.86
7	21	120.84	38.17	−10.79
8	27	154.70	49.74	−10.79
9	13	81.17	31.57	−9.64
10	19	114.69	40.81	−9.69
11	11	64.18	24.78	−9.84
12	20	119.91	42.14	−11.16
13	18	101.31	34.98	−10.85
14	26	157.29	53.38	−11.19
15	19	115.01	37.06	−10.32
16	28	165.52	49.04	−10.32
17	21	120.37	36.14	−10.71
18	21	120.95	39.71	−10.82
19	27	153.79	44.88	−10.66
20	27	154.73	50.05	−10.80

The definitions of the descriptors are given in Table 2.

[a] The numbers refer to the numbers of the molecules given in Table 1.

Table 4
Architecture and specification of the generated ANN

No. of nodes in the input layer	4
No. of nodes in the hidden layer	6
No. of nodes in the output layer	2
Weights learning rate	0.16
Biases learning rate	0.5
Momentum	0.5
Transfer function	Sigmoid

training was stopped at this point. Based upon the high values of iterations, two points may arise. First, the architecture of the generated ANN was correctly designed and second the descriptors appearing in the MLR models have been chosen adequately.

For the evaluation of the predictive power of the network, a trained ANN was used to predict the Kovats retention indices of the molecules included in the prediction set. Table 1 represents the experimental and ANN predicted values of retention indices on OV-1 and SE-30 stationary phases for the training and prediction set compounds. The statistical parameters obtained by ANN and MLR models for the training and prediction set compounds are shown in Table 5. It can be seen from this table that although the parameters appearing in the MLR models are used as inputs for the generated ANN, the statistics results of the latter show a large improvement. In the case of the ANN, the maximum relative error for predicted retention indices on OV-1 and SE-54 stationary phases are 6.61 and −6.45% for 3-pentanol and 2-methyl-3-pentanone, respectively, and the minimum values are −0.30 and −0.04% for 4-methyl-2-pentanone and 3,5-dimethyl-3-hexanol, respectively. The average percentage deviation between the predicted and the experimental values of Kovats retention indices for the prediction set are 2.5 and 3.0% on OV-1 and SE-54 stationary phases, respectively [30]. However, it is worth noting that these values are in agreement with the results obtained by experiments.

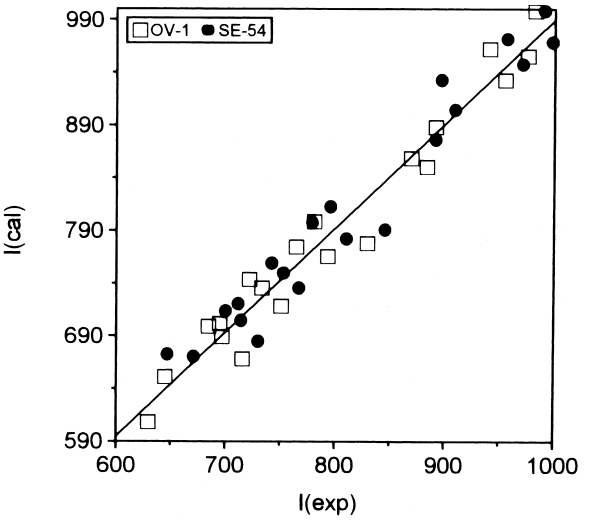

Fig. 1. Plot of the calculated retention indices against the experimental values.

Fig. 1 shows the plot of the ANN predicted versus the experimental values for the retention indices of the prediction set. The residuals of the ANN calculated values of the retention indices are plotted against the experimental values in Fig. 2. The propagation of the residuals on both sides of the zero line indicates that no systematic error exists in the development of the ANN.

4. Conclusion

The results of this study demonstrate that the QSPR method using the ANN techniques can generate a suitable model for the prediction of retention

Table 5
Statistical parameters obtained using the ANN and MLR models

Column	Model	R_p	R_t	SEC (%)	SEP (%)	F_t	F_p
OV-1	ANN	0.977	0.971	25.54	24.47	1419	381
	MLR	0.959	0.936	37.68	32.37	605	210
SE-54	ANN	0.975	0.970	25.74	25.49	1381	343
	MLR	0.955	0.934	38.24	34.02	579	185

Subscript t is referring to the training set, p is referring to the prediction set, R is the correlation coefficient, F is the statistical F-value.

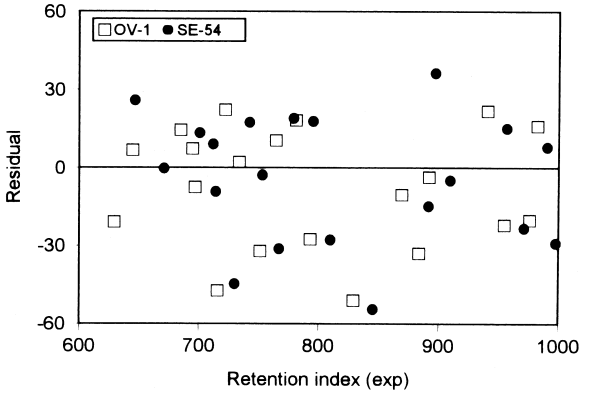

Fig. 2. Plot of the residuals versus experimental retention indices.

indices on SE-54 and OV-1 stationary phases simultaneously. The parameters of number of atoms in each molecule, the molecular volume, molecular shadow area on the *xy* plane and energy level of the highest occupied molecular orbital can be considered as comprehensive descriptors for predicting the partition coefficient of oxygen-containing compounds on nonpolar stationary phases. Finally, descriptors that appear in the MLR models and are included in the ANN provide information related to the different molecular properties that can participate in the physiochemical process that occurs in gas chromatographic separation.

Acknowledgements

This work was supported by the research deputy of Mazandaran University. The author also acknowledges M. Haghdadi for helpful discussion.

References

[1] L.S. Anker, P.C. Jurs, P.A. Edwards, Anal. Chem. 62 (1990) 2676.
[2] R. Kaliszan, Structure and Retention in Chromatography, Harwood, Amsterdam, 1997.
[3] O. Mekenyan, N. Dimov, V. Enchev, Anal. Chim. Acta 260 (1992) 69.
[4] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 64 (1992) 3059.
[5] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 65 (1993) 582.
[6] E.R. Collantes, W. Tong, W.J. Welsh, W.L. Zielinski, Anal. Chem. 68 (1996) 2083.
[7] A.R. Katritzky, K. Chen, U.R. Maran, D.A. Carison, Anal. Chem. 72 (2000) 101.
[8] P. Payares, D. Diaz, J. Olivero, R. Vivas, I. Gomez, J. Chromatogr. A 771 (1997) 213.
[9] J.M. Sutter, T.A. Pterson, P.C. Jurs, Anal. Chim. Acta 342 (1997) 113.
[10] M. Pompe, M. Novic, J. Chem. Inf. Comput. Sci. 39 (1999) 59.
[11] J. Kang, C. Cao, Z. Li, J. Chromatogr. A 799 (1998) 361.
[12] S.L. Anker, P.C. Jurs, Anal. Chem. 64 (1992) 1157.
[13] W.L. Xing, X.W. He, Anal. Chim. Acta 349 (1997) 283.
[14] M. Jalali-Heravi, M.H. Fatemi, Anal. Chim. Acta 415 (2000) 95.
[15] E. Maremgo, M.C. Gennaro, S. Anglino, J. Chromatogr. A 799 (1998) 47.
[16] G. Sacchero, M.C. Bruzzonoti, C. Sarzamini, E. Mentasti, H.J. Metting, P.M.J. Coenegracht, J. Chromatogr. A 799 (1998) 35.
[17] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 915 (2001) 177.
[18] K.L. Peterson, Anal. Chem. 64 (1992) 379.
[19] M. Jalali-Heravi, F. Parastar, J. Chem. Inf. Comput. Sci. 40 (2000) 147.
[20] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 825 (1998) 283.
[21] M. Jalali-Heravi, M.H. Fatemi, J. Chromatogr. A 897 (2000) 227.
[22] X. Zhang, P. Lu, J. Chromatogr. A 731 (1996) 187.
[23] E.K. Whalden, P.C. Jurs, Anal. Chem. 53 (1981) 484.
[24] T.R. Stouch, P.C. Jurs, J. Chem. Inf. Comput. Sci. 26 (1986) 26.
[25] J.J.P. Stewart, MOPAC, Semiempirical Molecular Orbital Program, QCPE, 455 (1983), version 6, 1990.
[26] S. Haykin, Neural Network, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[27] M.T. Beal, H.B. Hagan, M. Demuth, Neural Network Design, PWS, Boston, MA, 1996.
[28] N.K. Bose, P. Liang, Neural Network, Fundamentals, McGraw-Hill, New York, 1996.
[29] T.B. Blank, S.T. Brown, Anal. Chem. 65 (1993) 3084.
[30] A. Jouyban-Gharamaleki, M.G. Khaledi, B.J. Clark, J. Chromatogr. A 868 (2000) 277.