DOI:10.7598/cst2018.1501
ISSN:2278-3458

Chemical Science Transactions
2018, **7(4),** 558-575

RESEARCH ARTICLE

QSRR Study of Linear Retention Indices for Volatile Compounds using Statistical Methods

ASSIA BELHASSAN[1,2], SAMIR CHTITA[1],
TAHAR LAKHLIFI[1] and MOHAMMED BOUACHRINE[2*]

[1]MCNS Laboratory, Faculty of Science, Moulay Ismail University, Meknes, Morocco
[2]Materials, Environment and Modeling Laboratory, (ESTM) High School of Technology, Moulay Ismail University, Meknes, Morocco
m.bouachrine@est-umi.ac.ma

Received 7 January 2018 / Accepted 30 January 2018

Abstract: ACD/ChemSketch, MarvinSketch and ChemOffice programs were used to calculate several molecular descriptors of 138 volatile compounds (32 hydrocarbons, 29 ketones, 28 aldehydes, 23 alcohols, 7 carboxylic acids, 6 halogenated compounds, 4 furans, 2 pyrazines, 1 ester, 1 sulphur compounds, 1 pyridine, 1 amine and three other compounds). The best descriptors were selected to establish the quantitative structure retention relationship (QSRR) of linear retention indices of volatile compounds using multiple linear regressions (MLR), multiple non-linear regressions (MNLR) and artificial neural network (ANN) methods. We propose quantitative models according to these analyses. The models were used to predict the linear retention indices of the test set compounds and agreement between the experimental and predicted values was verified. The descriptors showed by QSRR study were used for study and designing of new compounds. The statistical results indicate that the predicted values are in good agreement with the experimental results. To validate the predictive power of the resulting models, external validation multiple correlation coefficient was calculated and has both in addition to a performance prediction power, a favorable estimation of stability.

Keywords: Volatile compounds, Linear retention indices, Quantitative structure retention relationship, Multiple linear regression, Artificial neural network

Introduction

Retention is a phenomenon primarily dependent on the interactions between the solute and the stationary phase of molecules, which included directional force, induction force, dispersion force and hydrogen bond[1]. These forces can be related to the structure of molecules; therefore, it was possible to predict the solute retention from molecular descriptors. Quantitative structure retention relationships (QSRR) studies have received much attention in chemometrics, biological chemistry, medicinal chemistry and many other fields. QSRR models are mathematical equations relating chemical structure to their property.

A number of reports that deals with QSRR retention calculation of several compounds have been published in the literature[2,3]. QSRR has been demonstrated to be a powerful tool for the investigation of the chromatographic parameters. In QSRR approaches, the structural features of solutes encoded by chemical descriptors and then the relationship between these descriptors and solute retention were studied[4].

The QSRR models can be applied to multiple linear regressions (MLR). Because of the complexity of relationships between the property of molecules and structures, nonlinear models are also used to model the structure–property relationships. Multiple non-linear regression (MNLR) and artificial neural network (ANN) methods are nonparametric nonlinear modeling technique that have attracted increasing interest.

The purpose of this work is to study the linear retention indices (LRI) of volatile compounds using QSRR chemical modeling methods. The main steps in this method includes: data collection, molecular descriptors generation, descriptor selection, model development and finally model performance evaluation[5].

Experimental

In this study, we selected 138 volatile compounds with linear retention indices (LRI) reported in the literature[6] to provide a diversified set of chemical families (hydrocarbons, alcohols, aldehydes, ketones, esters *etc*. The list of molecules and the log(LRI) values are listed in Table 1.

A total of 97 molecules were placed in the training set to build the QSRR models, whereas the remaining, 41 molecules constitute the test set. The division was carried out by random selection using the SPSS 19.0 statistical package[7].

Molecular descriptors

A wide variety of molecular descriptors were calculated using ACD/ChemSketch, MarvinSketch and ChemOffice software[8-10] to predict the correlation between these descriptors and the linear retention indices of the molecules studied (Table 2). Tables S1 show values of these descriptors for each molecule studied.

Statistical analysis

Multiple linear regression (MLR) is used to study the relationship between a dependent variable and several independent variables; it minimizes the differences between actual and predicted values and has been used to select the descriptors to be used as inputs in multiple non-linear regression (MNLR) and artificial neural network ANN. Linear and non-linear approaches were used to predict the effects on the linear retention indices, the equations were justified by the correlation coefficient (r), the mean square error (MSE), the fisher value (F) and the significance level (p)[11]. MLR, MNLR and ANN are generated using the SPSS 19.0 statistical package[7].

Cross-validations, the most commonly used techniques for internal validation, are statistical techniques in which different proportions of chemicals are iteratively held-out from the training set used for model development (an optimal parameters K selection step) and “predicted” as new by the developed model in order to verify internal “predictivity”. In this work the Leave-One-Out is used, this procedure successively removes one molecule from the training set containing 97 molecules. A QSRR model is constructed on a "96" set of compounds and the molecule removed is predicted by the model. This procedure is repeated "97" times in order to predict the property of all molecules[12].

Table 1. List of aroma compounds

N°	Molecule name	Log(LRI)	N°	Molecule name	Log(LRI)
1	Octane	2.904	42[a]	2,3-Butanedione	2.990
2	Nonane	2.954	43	2,5-Octadione	3.124
3	Pentadecane	3.175	44	(Z,E)-3,5-Octadien-2-one	3.187
4	Heptadecane	3.066	45	(E,E)-3,5-Octadien-2-one	3.202
5	2-Methylheptane	2.875	46[a]	3,5-Dimethyl-2,7-octadione	3.277
6[a]	4-Methylheptane	2.879	47	6-Methyl-3,5-heptadien-2-one	3.208
7	4-Methyloctane	2.931	48	6-Methyl-5-hepten-2-one	3.131
8	2-Methylnonane	2.979	49[a]	1-Hydroxy-2-propanone	3.120
9	4-Methyldecane	3.023	50	3-Hydroxy-2-butanone	3.115
10	5-Methyldecane	3.021	51	2,2,6-Trimethylcyclohexanone	3.128
11	2-Methylundecane	3.063	52[a]	3,5,5-Trimethyl-2-cyclohexen-1-one	3.156
12	3-Methylundecane	3.092	53	2-Propyl-2-cyclohexenone	3.170
13	4-Ethyldecane	3.014	54	γ-Valerolactone	3.216
14	2,4-Dimethylhexane	2.975	55	γ-Butyrolactone	3.222
15[a]	2,4-Dimethylheptane	2.906	56	1-Phenylethanone	3.226
16[a]	2,7-Dimethyloctane	2.960	57	γ-Hexalactone	3.240
17	2,5-Dimethylnonane	3.000	58[a]	5-Ethyl-2(5H)-furanone	3.254
18	3,7-Dimethyldecane	3.034	59	4-Ketoisophorone	3.237
19[a]	4,4-Dimethylundecane	2.998	60	α-Ionone	3.276
20	2,4-Dimethylundecane	3.076	61	β-Ionone	3.296
21	2,6-Dimethylundecane	3.077	62	Ethanal	2.848
22[a]	4,8-Dimethylundecane	3.082	63	Propanal	2.902
23	4,6-Dimethyldodecane	3.088	64[a]	Butanal	2.946
24	2,4,6-Trimethyloctane	3.085	65	Pentanal	2.994
25	(E)-2-Octene	2.940	66	Hexanal	3.040
26[a]	8-Heptadecene	3.237	67	Heptanal	3.078
27	2,4-Dimethyl-1-heptene	2.946	68	Octanal	3.115
28	1,4-Octadiene	3.268	69	Nonanal	3.149
29	1,3-Cyclooctadiene	3.229	70	Decanal	3.180
30[a]	3-Ethyl-2-methyl-1,3-hexadiene	3.160	71[a]	(E)-2-Butenal	3.023
31	1,3-Dimethylbenzene	3.065	72	(E)-2-Pentenal	3.060
32	1,2,3-Trimethylbenzene	3.120	73	(E)-2-Hexenal	3.092
33	2-Propanone	2.913	74	(E)-2-Heptenal	3.128
34	2-Butanone	2.960	75	(Z)-4-Heptenal	3.099
35[a]	2-Pentanone	2.994	76[a]	(E)-2-Octenal	3.161
36[a]	3-Hexanone	3.027	77[a]	(Z,E)-2,4-Heptadienal	3.173
37[a]	2-Heptanone	3.077	78[a]	(E,E)-2,4-Heptadienal	3.181
38	2-Octanone	3.113	79	(E,E)-2,6-Nonadienal	3.207
39	3-Penten-2-one	3.058	80	2-Methylbutanal	2.964
40[a]	1-Octen-3-one	3.119	81	3-Methylbutanal	2.966
41	3-Octen-2-one	3.131	82[a]	2-Methyl-2-propenal	2.947

Contd…

N°	Molecule name	Log(LRI)	N°	Molecule name	Log(LRI)
83[a]	(E)-2-Methyl-2-butenal	3.046	111[a]	1,2-Dimethylcyclohexanol	3.214
84	2-Methyl-2-pentenal	3.072	112	Benzenemethanol	3.279
85	Benzaldehyde	3.191	113[a]	Ethanoic acid	3.164
86[a]	4-Ethylbenzaldehyde	3.241	114[a]	Butanoic acid	3.215
87[a]	4-Propylbenzaldehyde	3.272	115	Hexanoic acid	3.274
88	β-Cyclocitral	3.220	116	Heptanoic acid	3.298
89	Safranal	3.226	117	2-Propenoic acid	3.216
90	1-Pentanol	3.100	118	2-Methylpropanoic acid	3.198
91[a]	3-Pentanol	2.957	119	3-Methylbutanoic acid	3.230
92	1-Hexanol	3.133	120[a]	Lactic acid ethyl ester	3.131
93	1-Octanol	3.194	121	Iodo-methane	2.910
94	1-Penten-3-ol	3.068	122[a]	Iodo-ethane	2.947
95	(E)-2-Penten-1-ol	3.120	123	2-Iodo-propane	2.960
96[a]	(Z)-2-Penten-1-ol	3.123	124	1-Iodo-pentane	3.072
97	1-Octen-3-ol	3.162	125	1-Iodo-octane	3.178
98	2-Octen-1-ol	3.210	126	5-Chloro-1-pentene	3.012
99	2,3-Butanediol	3.187	127[a]	Dimethyl sulphide	2.876
100[a]	(Z)-1,5-Octadien-3-ol	3.173	128	2-Methylfuran	2.942
101	1,7-Octadien-3-ol	3.179	129	2-Ethylfuran	2.981
102	3,5-Octadien-2-ol	3.154	130[a]	2-Pentylfuran	3.095
103	2-Methyl-3-pentanol	3.134	131	cis-2-(2-Pentenyl)furan	3.118
104[a]	(E)-2-Ethyl-1-hexanol	3.251	132	2.6-Dimethylpyrazine	3.129
105	(Z)-2-Ethyl-1-hexanol	3.175	133[a]	2,3,5,6-Tetramethylpyrazine	3.176
106	2-(2-Methoxypropoxy)-1-propanol	3.173	134[a]	2-Ethylpyridine	3.116
107	1-(2-Methoxypropoxy)-2-propanol	3.187	135	N,N-Dimethyl-methanamine	2.796
108	1-(2-Methoxy-1-methyl-ethoxy)-2-propanol	3.181	136[a]	1-Acetyl-2,2-dimethylcyclobutane	3.006
109[a]	1-Ethoxy-2-propanol	3.069	137	1-Methoxy-4-(1′-methylethyl)cyclohexa-1,4-diene	3.300
110[a]	Cyclooctanol	3.209	138	2,7-Epoxy-megastigma-4,8-diene	3.178

[a]*Test Set*

Y-randomization, randomly scrambling the responses, is another internal validation approach that must be used in parallel with cross-validations, and must always be applied to test the significance of the derived QSRR model, highlighting the presence of apparent models, obtained only by chance correlation[12]. We performed in this work 100-y-randomization tests for the MLR and MNLR models. In this test, random QSPR models are generated by randomly shuffling the dependent variable while keeping the independent variables as it is. The new QSRR models are expected to have significantly low r^2 and r_{CV}^2 values for several trials, which confirm that the developed QSRR models are robust.

Table 2. Calculated chemical descriptors

Software	Descriptors	Abbreviation
ChemOffice	Heat of formation, kJ mol^{-1}	$H°$
	Gibbs free energy, kJ mol^{-1}	G
	Ideal gas thermal capacity, J mol^{-1} K^{-1}	IGTC
	Melting point(Kelvin)	T
	Critical temperature, K	CT
	Boiling point(Kelvin)	TB
	Critical pressure(Bar)	CP
	Henry's law constant	KH
	Total valence connectivity	TVC
	Partition coefficient	PC
	Number of rotatable bonds	NRB
	Shape coefficient	I
	Sum of valence degrees	SVD
	Total connectivity	TC
ChemSketch	Percent ratios of hydrogen	H%
	Percent ratios of oxygen	O%
	Percent ratios of carbon	C%
	Surface tension	γ
	Index of refraction	n
	Density	d
MarvinSketch	Log P	Log P
	Winner index	W
	Number of H-Bond acceptors	NHA
	Number of H-Bond donors	NHD
	Balaban index	J
	Polar surface area ($A°$)	PSA

Results and Discussion

Data set for analysis

A QSRR study was carried out for a series of 138 volatile compounds, as indicated above, to determine a quantitative relationship between the structure and the linear retention indices for molecules studied. The values of the 27 descriptors are shown in Table S1.

Multiple linear regressions (MLR)

In order to propose a mathematical model and to evaluate quantitatively the substituent's physicochemical effects on the log(LRI) of the totality of the set of these 138 molecules, we submitted the data matrix constituted obviously from the 27 variables corresponding to the 97 molecules (training set), to a stepwise multiple regression analysis.

Table S1. Chemical descriptors computed by ChemOffice, ChemSketch and MarvinSketch software

N°	*H*°	G	IGTC	T	CT	TB	CP	*K*H	TVC	PC	MTI	NRB	I	SVD	TC	H%	O%	C%	n	γ	D	log P	W	NHA	NHD	J	PSA
1	-208.45	16.48	189.224	179.42	561.030	393.054	25.354	-2.117	0.125	4.926	728	5	0	14	0.125	0.159	0	0.841	1.402	22.6	0.711	4.000	84	0	0	2.53	0
2	-229.09	24.90	212.154	190.69	583.165	415.868	23.070	-2.214	0.088	5.455	1024	6	1	16	0.088	0.157	0	0.843	1.408	23.5	0.724	4.510	120	0	0	2.60	0
3	-352.93	75.42	349.737	258.31	690.251	533.766	14.154	-2.952	0.011	8.629	4592	12	1	28	0.011	0.152	0	0.848	1.431	26.9	0.769	7.190	560	0	0	2.81	0
4	-394.21	92.26	395.597	280.85	718.397	565.834	12.311	-3.198	0.006	9.687	6656	14	1	32	0.006	0.151	0	0.849	1.435	27.5	0.777	8.102	816	0	0	2.85	0
5	-213.73	14.04	189.738	164.42	544.689	378.865	25.559	-2.091	0.144	4.796	688	4	1	14	0.144	0.159	0	0.841	1.400	21.8	0.709	3.746	79	0	0	2.72	0
6[a]	-213.73	14.04	189.738	164.42	544.689	378.865	25.559	-2.091	0.144	4.796	656	4	1	14	0.144	0.159	0	0.841	1.400	21.8	0.709	3.746	75	0	0	2.92	0
7	-234.37	22.46	212.668	175.69	567.818	402.224	23.248	-2.214	0.102	5.325	928	5	0	16	0.102	0.157	0	0.843	1.407	22.7	0.722	4.202	108	0	0	2.95	0
8	-255.01	30.88	235.599	186.96	589.456	424.678	21.236	-2.337	0.072	5.854	1336	6	1	18	0.072	0.156	0	0.844	1.412	23.5	0.733	4.658	158	0	0	2.77	0
9	-275.65	39.30	258.529	198.23	609.741	446.229	19.475	-2.460	0.051	6.383	1696	7	0	20	0.051	0.155	0	0.845	1.417	24.2	0.742	5.114	202	0	0	2.97	0
10	-275.65	39.30	258.529	198.23	609.741	446.229	19.475	-2.460	0.051	6.383	1680	7	0	20	0.051	0.155	0	0.845	1.417	24.2	0.742	5.114	200	0	0	3.02	0
11	-296.29	47.72	281.460	209.50	628.792	466.875	17.924	-2.583	0.036	6.912	2304	8	1	22	0.036	0.154	0	0.846	1.421	24.8	0.750	5.570	277	0	0	2.82	0
12	-296.29	47.72	281.460	209.50	628.792	466.875	17.924	-2.583	0.036	6.912	2248	8	1	22	0.036	0.154	0	0.846	1.421	24.8	0.750	5.570	270	0	0	2.90	0
13	-296.29	47.72	281.460	209.50	628.792	466.875	17.924	-2.583	0.036	6.912	2088	8	0	22	0.036	0.154	0	0.846	1.421	24.8	0.750	5.570	250	0	0	3.19	0
14	-219.01	11.60	190.252	149.42	527.715	364.348	25.767	-2.091	0.167	4.666	624	3	0	14	0.167	0.159	0	0.841	1.399	20.9	0.707	3.494	71	0	0	3.10	0
15[a]	-239.65	20.02	213.183	160.69	551.880	388.251	23.428	-2.214	0.118	5.195	880	4	1	16	0.118	0.157	0	0.843	1.405	21.9	0.721	3.950	102	0	0	3.15	0
16[a]	-260.29	28.44	236.113	171.96	574.474	411.251	21.394	-2.337	0.083	5.724	1280	5	0	18	0.083	0.156	0	0.844	1.411	22.8	0.732	4.406	151	0	0	2.91	0
17	-280.93	36.86	259.044	183.23	595.645	433.346	19.613	-2.460	0.059	6.253	1616	6	1	20	0.059	0.155	0	0.845	1.416	23.5	0.741	4.862	192	0	0	3.15	0
18	-301.57	45.28	281.974	194.50	615.519	454.538	18.046	-2.583	0.042	6.782	2080	7	0	22	0.042	0.154	0	0.846	1.420	24.2	0.749	5.318	249	0	0	3.18	0
19[a]	-320.4	61.42	303.449	238.19	642.928	478.901	16.701	-2.706	0.031	7.311	2592	8	1	24	0.031	0.153	0	0.847	1.425	25.1	0.758	6.370	312	0	0	3.29	0
20	-322.21	53.70	304.904	205.77	634.212	474.825	16.660	-2.706	0.029	7.311	2704	8	1	24	0.029	0.153	0	0.847	1.423	24.7	0.755	5.774	326	0	0	3.12	0
21	-322.21	53.70	304.904	205.77	634.212	474.825	16.660	-2.706	0.029	7.311	2688	8	1	24	0.029	0.153	0	0.847	1.423	24.7	0.755	5.774	324	0	0	3.15	0
22[a]	-322.21	53.70	304.904	205.77	634.212	474.825	16.660	-2.706	0.029	7.311	2624	8	1	24	0.029	0.153	0	0.847	1.423	24.7	0.755	5.774	316	0	0	3.23	0
23	-342.85	62.12	327.835	217.04	651.824	494.209	15.427	-2.829	0.021	7.840	3240	9	0	26	0.021	0.152	0	0.848	1.426	25.2	0.761	6.230	392	0	0	3.30	0
24	-286.21	34.42	259.558	168.23	581.012	420.135	19.753	-2.460	0.068	6.123	1504	5	0	20	0.068	0.155	0	0.845	1.414	22.8	0.739	4.610	178	0	0	3.43	0
25	-91.23	96.70	174.794	174.34	578.067	400.177	26.735	-1.485	0.083	4.442	728	4	0	16	0.125	0.144	0	0.856	1.421	22.9	0.727	3.284	84	0	0	2.53	0
26[a]	-276.99	172.48	381.168	275.77	728.182	570.451	12.773	-2.592	0.004	9.203	6656	13	1	34	0.006	0.144	0	0.856	1.446	27.8	0.787	7.388	816	0	0	2.85	0
27	-118.73	101.75	202.191	159.97	564.440	394.206	24.338	-1.608	0.072	4.711	880	4	1	18	0.118	0.144	0	0.856	1.416	21.8	0.732	3.485	102	0	0	3.15	0
28	34.20	184.54	164.251	172.58	580.858	398.457	27.936	-1.157	0.048	3.958	728	4	0	18	0.125	0.128	0	0.872	1.434	22.9	0.738	3.039	84	0	0	2.53	0
29	-30.55	84.36	125.858	185.52	647.038	419.897	37.823	-0.784	0.028	3.564	576	0	0	20	0.063	0.112	0	0.888	1.473	30.9	0.841	2.760	64	0	0	2.00	0
30[a]	-6.02	175.86	187.313	155.93	591.722	408.834	25.559	-1.335	0.051	4.287	832	3	0	20	0.118	0.130	0	0.870	1.435	22.1	0.750	3.091	96	0	0	3.41	0
31	16.61	119.26	131.214	218.36	636.244	421.471	35.856	0.532	0.028	3.140	552	0	0	22	0.083	0.095	0	0.905	1.500	28.7	0.870	3.200	61	0	0	2.23	0
32	-15.50	118.05	156.006	242.15	653.701	443.144	31.526	0.759	0.024	3.539	728	0	0	24	0.068	0.101	0	0.899	1.500	28.7	0.869	3.111	82	0	0	2.15	0
33	-217.83	-154.54	74.943	173.00	494.136	317.977	48.025	2.980	0.204	-0.208	84	0	1	12	0.577	0.104	).275	0.620	1.345	18.8	0.772	-0.240	9	1	0	2.32	17.1
34	-238.47	-146.12	97.873	184.27	521.407	343.544	42.220	2.720	0.144	0.321	158	1	0	14	0.408	0.112	).222	0.666	1.367	21.0	0.786	0.275	18	1	0	2.54	17.1

Contd…

35[a]	-259.11	-137.70	120.804	195.54	546.672	368.208	37.408	2.495	0.102	0.850	270	2	1	16	0.289	0.117	).186	0.697	1.383	22.6	0.796	0.945	32	1	0	2.6317.1
36[a]	-279.75	-129.28	143.734	206.81	570.140	391.968	33.374	2.324	0.072	1.379	416	3	0	18	0.204	0.121	).160	0.720	1.394	23.9	0.803	1.401	50	1	0	2.8317.1
37[a]	-300.39	-120.86	166.665	218.08	591.988	414.823	29.959	2.210	0.051	1.908	640	4	1	20	0.144	0.124	).140	0.736	1.403	24.9	0.808	1.857	79	1	0	2.7217.1
38	-321.03	-112.44	189.595	229.35	612.372	436.775	27.042	2.100	0.036	2.437	914	5	0	22	0.102	0.126	).125	0.749	1.410	25.8	0.812	2.313	114	1	0	2.7217.1
39	-141.89	-57.48	106.374	190.46	565.427	375.623	39.906	3.062	0.068	0.516	270	1	1	18	0.289	0.096	).190	0.714	1.411	23.1	0.826	0.771	32	1	0	2.6317.1
40[a]	-195.60	-24.60	179.052	227.59	615.067	435.166	28.263	2.715	0.021	2.103	890	5	0	24	0.102	0.112	).127	0.761	1.422	25.8	0.825	2.608	110	1	0	2.8817.1
41	-203.81	-32.22	175.165	224.27	628.530	443.355	28.566	2.702	0.024	2.103	914	4	0	24	0.102	0.112	).127	0.761	1.431	26.3	0.833	2.139	114	1	0	2.7517.1
42[a]	-351.05	275.04	98.245	234.20	580.202	390.879	45.902	5.093	0.042	-1.368	224	1	0	22	0.333	0.070	).372	0.558	1.381	26.4	0.969	-1.340	29	2	0	2.9934.1
43	-433.61	241.36	189.966	279.28	657.834	477.047	28.905	5.496	0.010	0.905	1084	5	0	30	0.083	0.099	).225	0.676	1.419	29.4	0.918	0.628	143	2	0	3.1234.1
44	-86.59	48.00	160.736	219.19	645.194	449.848	30.223	3.302	0.016	1.679	914	3	0	26	0.102	0.097	).129	0.774	1.454	26.8	0.856	1.965	114	1	0	2.7517.1
45	-86.59	48.00	160.736	219.19	645.194	449.848	30.223	3.302	0.016	1.679	914	3	0	26	0.102	0.097	).129	0.774	1.454	26.8	0.856	1.965	114	1	0	2.7517.1
46[a]	-485.45	229.40	236.856	271.82	666.641	492.438	24.124	5.250	0.007	0.876	1684	5	0	34	0.056	0.107	).188	0.705	1.426	28.2	0.902	1.036	223	2	0	3.6134.1
47	-96.38	39.45	160.802	205.23	640.795	444.099	30.390	3.253	0.020	1.549	868	2	1	26	0.118	0.097	).129	0.774	1.452	25.7	0.857	1.493	108	1	0	2.9117.1
48	-213.60	-40.77	175.231	210.31	623.846	437.529	28.719	2.653	0.029	1.823	868	3	1	24	0.118	0.112	).127	0.761	1.430	25.2	0.835	1.127	108	1	0	2.9117.1
49[a]	-370.06	291.36	87.958	233.82	581.929	409.077	54.788	7.130	0.065	-0.733	138	1	0	18	0.408	0.082	).432	0.486	1.397	32.3	1.019	-1.132	18	2	1	2.5437.3
50	-395.98	285.38	111.403	230.09	577.980	410.748	48.292	7.007	0.053	-0.204	224	1	0	20	0.333	0.092	).363	0.545	1.408	30.4	0.983	-0.874	29	2	1	2.9937.3
51	-317.57	-86.44	180.519	285.95	684.788	467.908	29.861	2.310	0.021	2.422	878	0	0	26	0.059	0.115	).114	0.771	1.431	24.8	0.869	2.259	106	1	0	2.6017.1
52[a]	-250.92	-58.40	172.455	303.47	694.377	474.612	31.704	2.639	0.015	2.278	896	0	0	28	0.059	0.102	).116	0.782	1.455	26.4	0.905	1.865	109	1	0	2.5117.1
53	-245.82	-45.20	172.707	283.81	706.296	490.534	31.280	3.206	0.010	2.298	992	2	1	28	0.042	0.102	).116	0.782	1.469	30.9	0.931	1.773	121	1	0	2.2517.1
54	-391.01	258.40	114.075	205.88	697.621	464.744	40.934	2.254	0.024	-0.284	328	0	1	24	0.118	0.081	).320	0.600	1.430	29.3	1.040	0.932	41	1	0	1.9326.3
55	-350.03	259.11	88.959	198.85	689.010	450.107	48.630	2.377	0.029	-0.803	212	0	0	22	0.144	0.070	).372	0.558	1.442	35.4	1.128	-0.475	26	1	0	2.1826.3
56	-84.50	-0.03	129.725	255.77	684.789	462.990	39.456	3.368	0.007	1.581	726	1	0	30	0.059	0.067	).133	0.800	1.511	34.1	0.993	1.625	88	1	0	1.9817.1
57	-411.65	249.98	137.005	217.15	712.100	484.582	36.333	2.131	0.017	0.245	496	1	0	26	0.083	0.088	).280	0.631	1.431	29.0	1.002	1.388	63	1	0	2.2226.3
58[a]	-353.87	220.02	129.267	217.91	718.649	487.120	37.963	2.460	0.011	0.381	496	1	0	28	0.083	0.072	).285	0.643	1.455	30.5	1.046	1.214	63	1	0	2.2226.3
59	-388.62	180.99	177.402	371.69	756.520	526.996	32.992	6.259	0.004	1.029	1066	0	0	36	0.048	0.079	).210	0.710	1.469	30.8	1.020	0.936	138	2	0	2.4334.1
60	-211.48	41.46	247.358	320.94	734.941	532.652	22.398	3.187	0.003	3.710	2400	2	0	38	0.017	0.105	).083	0.812	1.511	32.7	0.935	2.751	301	1	0	2.5217.1
61	-202.61	39.54	247.033	337.70	737.307	536.100	22.784	3.187	0.003	3.770	2400	2	0	38	0.017	0.105	).083	0.812	1.517	33.6	0.944	2.597	301	1	0	2.5217.1
62	-170.19	133.56	55.999	153.80	486.186	307.015	56.026	2.629	0.236	-0.224	40	0	1	10	0.707	0.092	).363	0.545	1.314	17.6	0.748	0.054	4	1	0	1.6317.1
63	-190.83	125.14	78.930	165.07	514.311	332.961	48.766	2.507	0.167	0.305	90	1	0	12	0.500	0.104	).275	0.620	1.347	20.5	0.770	0.590	10	1	0	1.9717.1
64[a]	-211.47	116.72	101.860	176.34	540.305	358.002	42.831	2.272	0.118	0.834	170	2	1	14	0.354	0.112	).222	0.666	1.369	22.5	0.784	0.966	20	1	0	2.1917.1
65	-232.11	-108.3	124.790	187.61	564.395	382.140	37.916	2.221	0.083	1.363	288	3	0	16	0.250	0.117	).186	0.697	1.384	24.0	0.794	1.422	35	1	0	2.3417.1
66	-252.75	-99.88	147.721	198.88	586.774	405.374	33.802	2.060	0.059	1.892	452	4	1	18	0.177	0.121	).160	0.720	1.396	25.1	0.801	1.878	56	1	0	2.4517.1
67	-273.39	-91.46	170.651	210.15	607.611	427.704	30.323	1.859	0.042	2.421	670	5	0	20	0.125	0.124	).140	0.736	1.405	26.0	0.807	2.334	84	1	0	2.5317.1
68	-294.03	-83.04	193.582	221.42	627.054	449.130	27.354	1.678	0.029	2.950	950	6	1	22	0.088	0.126	).125	0.749	1.412	26.8	0.811	2.790	120	1	0	2.6017.1
69	-314.67	-74.62	216.512	232.69	645.230	469.652	24.801	1.523	0.021	3.479	1300	7	0	24	0.063	0.128	).112	0.760	1.417	27.4	0.815	3.246	165	1	0	2.6517.1
70	-335.31	-66.20	239.442	243.96	662.251	489.270	22.590	1.581	0.015	4.008	1728	8	1	26	0.044	0.129	).102	0.769	1.422	27.9	0.818	3.702	220	1	0	2.6917.1
71[a]	-94.25	-36.50	87.430	171.26	559.755	365.533	45.902	3.066	0.079	0.520	170	1	1	16	0.354	0.086	).228	0.685	1.402	23.0	0.819	0.792	20	1	0	2.1917.1

Contd...

72	-114.89	-28.08	110.361	182.53	582.954	389.393	40.466	2.781	0.056	1.049	288	2	0	18	0.250	0.096	).1900.714	1.413	24.6	0.824	1.248	35	1	0	2.3417.1
73	-135.53	-19.66	133.291	193.8	604.447	412.348	35.942	2.699	0.039	1.578	452	3	1	20	0.177	0.103	).1630.734	1.421	25.7	0.828	1.704	56	1	0	2.4517.1
74	-156.17	-11.24	156.221	205.07	624.405	434.400	32.137	2.541	0.028	2.107	670	4	0	22	0.125	0.108	).1430.750	1.427	26.6	0.830	2.160	84	1	0	2.5317.1
75	-156.17	-11.24	156.221	205.07	624.405	434.400	32.137	2.541	0.028	1.937	670	4	0	22	0.125	0.108	).1430.750	1.427	26.6	0.830	1.620	84	1	0	2.5317.1
76[a]	-176.81	-2.82	179.152	216.34	642.975	455.547	28.905	2.523	0.020	2.636	950	5	1	24	0.088	0.112	).1270.761	1.432	27.3	0.832	2.616	120	1	0	2.6017.1
77[a]	-38.95	68.98	141.792	199.99	641.743	441.010	34.119	3.141	0.019	1.683	670	3	0	24	0.125	0.092	).1450.763	1.453	27.2	0.856	1.986	84	1	0	2.5317.1
78[a]	-38.95	68.98	141.792	199.99	641.743	441.010	34.119	3.141	0.019	1.683	670	3	0	24	0.125	0.092	).1450.763	1.453	27.2	0.856	1.986	84	1	0	2.3517.1
79	-80.23	85.82	187.652	222.53	675.804	481.844	27.585	2.901	0.009	2.681	1300	5	0	28	0.063	0.102	).1160.782	1.457	28.5	0.855	2.358	165	1	0	2.6517.1
80	-237.39	110.74	125.305	172.61	547.521	367.698	38.293	2.145	0.096	1.363	260	2	1	16	0.289	0.117	).1860.697	1.382	22.6	0.791	1.170	31	1	0	2.7517.1
81	-237.39	110.74	125.305	172.61	547.521	367.698	38.293	2.132	0.096	1.233	266	2	1	16	0.289	0.117	).1860.697	1.382	22.6	0.791	1.170	32	1	0	2.6317.1
82[a]	-95.83	-37.43	91.383	160.62	536.918	349.399	45.592	2.852	0.083	0.300	156	1	0	16	0.408	0.086	).2280.685	1.384	20.8	0.807	0.789	18	1	0	2.5417.1
83[a]	-124.68	-36.63	110.427	168.57	577.373	382.971	40.725	2.745	0.068	0.829	260	1	1	18	0.289	0.096	).1900.714	1.411	23.1	0.826	0.776	31	1	0	2.7517.1
84	-145.32	-28.21	133.357	179.84	599.295	406.173	36.159	2.625	0.048	1.358	410	2	0	20	0.204	0.103	).1630.734	1.419	24.4	0.829	1.232	50	1	0	2.8317.1
85	-36.86	20.95	110.781	236.57	683.038	454.392	45.347	2.959	0.008	1.495	530	1	0	28	0.072	0.057	).1510.792	1.566	38.8	1.049	1.455	64	1	0	2.1317.1
86[a]	-89.61	28.16	158.503	271.63	711.743	494.068	34.930	2.684	0.005	2.523	1008	2	0	32	0.042	0.075	).1190.806	1.548	36.3	1.001	2.214	125	1	0	2.1717.1
87[a]	-110.25	36.58	181.433	282.90	724.923	512.555	31.280	2.564	0.003	3.052	1360	3	1	34	0.029	0.082	).1080.810	1.539	36.0	0.985	2.670	170	1	0	1.9417.1
88	-230.91	-36.54	196.658	301.04	694.045	486.905	29.282	2.666	0.010	3.350	1138	1	0	30	0.042	0.106	).1050.789	1.507	33.3	0.947	2.326	140	1	0	2.4017.1
89	-173.13	-6.58	188.920	301.80	700.094	489.428	30.457	3.266	0.007	2.926	1138	1	0	32	0.042	0.094	).1070.800	1.529	34.7	0.975	1.822	140	1	0	2.4017.1
90	-298.76	145.60	133.448	206.43	573.422	410.132	38.771	3.275	0.112	1.352	288	3	0	14	0.250	0.137	).1810.681	1.407	27.1	0.811	1.480	35	1	1	2.3420.2
91[a]	-304.04	148.04	133.962	191.43	547.228	388.822	39.160	3.266	0.129	1.132	264	2	1	14	0.289	0.137	).1810.681	1.405	25.6	0.809	1.210	31	1	1	2.7520.2
92	-319.40	137.18	156.378	217.70	594.658	432.273	34.521	3.113	0.079	1.881	452	4	1	16	0.177	0.138	).1570.705	1.415	27.9	0.816	2.025	56	1	1	2.4520.2
93	-360.68	120.34	202.239	240.24	633.340	473.841	27.877	3.002	0.040	2.939	950	6	1	20	0.088	0.139	).1230.738	1.427	29.0	0.823	3.150	120	1	1	2.6020.2
94	-178.61	-60.20	123.419	189.67	549.256	387.069	41.303	3.775	0.075	0.848	264	2	1	16	0.289	0.117	).1860.697	1.421	25.6	0.827	1.022	31	1	1	2.7520.2
95	-181.54	-65.38	119.018	201.35	589.514	417.048	41.409	3.811	0.075	1.068	288	2	0	16	0.250	0.117	).1860.697	1.437	27.9	0.842	0.751	35	1	1	2.3420.2
96[a]	-181.54	-65.38	119.018	201.35	589.514	417.048	41.409	3.811	0.075	1.068	288	2	0	16	0.250	0.117	).1860.697	1.437	27.9	0.842	0.751	35	1	1	2.3420.2
97	-240.53	-34.94	192.210	223.48	613.269	453.492	29.409	3.415	0.026	2.435	890	5	0	22	0.102	0.126	).1250.749	1.438	28.0	0.834	2.390	110	1	1	2.8820.2
98	-243.46	-40.12	187.809	235.16	647.030	479.922	29.473	3.451	0.026	2.655	950	5	1	22	0.088	0.126	).1250.749	1.449	29.7	0.845	2.119	120	1	1	2.6020.2
99	-440.91	295.72	124.561	225.98	579.345	429.939	50.875	5.026	0.067	-0.751	224	1	0	18	0.333	0.112	).3550.533	1.434	34.8	0.997	-0.920	29	2	2	2.9940.5
100[a]	-123.31	45.28	177.780	218.40	627.656	459.851	31.140	4.015	0.018	1.951	890	4	0	24	0.102	0.112	).1270.761	1.461	28.7	0.857	1.676	110	1	1	2.8820.2
101	-115.10	52.90	181.666	221.72	615.302	451.936	30.796	4.015	0.015	1.951	890	5	0	24	0.102	0.112	).1270.761	1.451	28.1	0.848	2.145	110	1	1	2.8820.2
102	-131.52	37.66	173.894	215.08	639.909	467.634	31.491	4.002	0.020	2.011	914	3	0	24	0.102	0.112	).1270.761	1.470	29.2	0.866	1.747	114	1	1	2.7520.2
103	-329.96	142.06	157.407	187.70	554.429	398.057	35.179	2.837	0.105	1.531	386	2	1	16	0.236	0.138	).1570.705	1.411	25.3	0.811	1.471	46	1	1	3.1420.2
104[a]	-365.96	122.78	202.753	225.24	620.319	461.693	28.114	2.897	0.046	2.809	838	5	1	20	0.102	0.139	).1230.738	1.426	28.0	0.821	2.581	104	1	1	3.0920.2
105	-365.96	122.78	202.753	225.24	620.319	461.693	28.114	2.897	0.046	2.809	838	5	1	20	0.102	0.139	).1230.738	1.426	28.0	0.821	2.581	104	1	1	3.0920.2
106	-615.04	343.64	210.480	243.43	629.885	471.289	30.323	7.329	0.012	-0.164	1016	5	0	30	0.083	0.109	).3240.567	1.422	29.5	0.955	0.053	141	3	1	3.1738.7
107	-615.04	343.64	210.480	243.43	621.182	464.778	30.323	7.329	0.012	-0.164	1054	5	0	30	0.083	0.109	).3240.567	1.422	29.5	0.955	0.053	146	3	1	3.0338.7

Contd...

108	-615.04	343.64	210.480	243.43	621.182	464.778	30.323	7.329	0.012	-0.164	1032	5	0	30	0.083	0.109	).324	0.567	1.422	29.5	0.955	0.053	143	3	1	3.12	38.7
109[a]	-436.26	253.04	149.034	213.66	572.202	412.673	38.340	5.520	0.053	0.092	400	3	0	20	0.204	0.116	).307	0.577	1.408	27.8	0.903	0.273	52	2	1	2.68	29.5
110[a]	-318.68	120.09	156.535	240.58	675.486	479.865	37.226	3.452	0.023	2.385	726	0	0	22	0.051	0.126	).125	0.749	1.466	30.7	0.922	2.285	88	1	1	1.89	20.2
111[a]	-311.46	109.09	165.657	267.28	632.221	450.116	35.558	3.452	0.032	2.305	670	0	0	22	0.072	0.126	).125	0.749	1.456	29.2	0.903	1.819	80	1	1	2.21	20.2
112	-103.51	-16.35	119.438	255.39	679.076	478.828	46.466	5.377	0.010	1.104	530	1	0	26	0.072	0.075	).148	0.777	1.546	40.7	1.047	1.018	64	1	1	2.13	20.2
113[a]	-434.88	377.95	65.634	272.40	594.481	395.480	57.305	4.720	0.091	-0.194	72	0	1	16	0.577	0.067	).533	0.400	1.375	31.9	1.068	-0.260	9	1	1	2.32	37.3
114[a]	-476.16	361.11	111.495	294.94	634.708	440.014	43.684	5.004	0.046	0.864	244	2	1	20	0.289	0.092	).363	0.545	1.411	32.5	0.987	0.885	32	1	1	2.63	37.3
115	-517.44	344.27	157.356	317.48	669.451	480.932	34.399	4.422	0.023	1.922	592	4	1	24	0.144	0.104	).275	0.620	1.427	32.8	0.950	1.912	79	1	1	2.72	37.3
116	-538.08	335.85	180.286	328.75	685.075	500.036	30.830	4.302	0.016	2.451	852	5	0	26	0.102	0.108	).246	0.646	1.432	32.9	0.938	2.214	114	1	1	2.72	37.3
117	-330.09	281.69	78.021	281.91	618.906	416.532	52.893	5.382	0.037	0.351	140	1	0	20	0.408	0.056	).444	0.500	1.422	32.8	1.063	0.685	18	1	1	2.54	37.3
118	-481.44	363.55	112.009	279.94	620.536	426.972	44.150	4.619	0.053	0.644	224	1	0	20	0.333	0.092	).363	0.545	1.408	30.4	0.983	0.594	29	1	1	2.99	37.3
119	-502.08	355.13	134.940	291.21	639.447	448.428	39.014	4.493	0.037	1.263	364	2	1	22	0.236	0.099	).313	0.588	1.418	30.9	0.962	1.050	48	1	1	2.95	37.3
120[a]	-621.32	458.41	149.875	233.76	602.937	439.370	39.407	6.335	0.015	0.331	504	3	0	28	0.167	0.085	).406	0.508	1.420	33.2	1.050	-0.217	70	2	1	3.17	46.5
121	12.90	15.66	45.881	158.59	558.579	337.232	56.026	0.659	1.435	1.466	10	0	0	1.486	1.000	0.021	0	0.085	1.527	27.2	2.222	1.600	1	0	0	1.00	0
122[a]	-7.74	24.08	68.811	169.86	583.293	362.122	48.766	0.549	1.015	1.995	36	0	1	3.486	0.707	0.032	0	0.154	1.516	28.5	1.940	2.000	4	0	0	1.63	0
123	-33.66	30.06	92.256	166.13	589.194	371.757	43.283	0.329	0.828	2.524	78	0	1	5.486	0.577	0.042	0	0.212	1.505	27.5	1.747	2.093	9	0	0	2.32	0
124	-69.66	49.34	137.603	203.67	645.943	431.368	33.802	0.161	0.359	3.582	272	3	0	9.486	0.250	0.056	0	0.303	1.499	30.4	1.524	3.348	35	0	0	2.34	0
125	-131.58	74.60	206.394	237.48	695.229	492.478	24.801	-0.199	0.127	5.169	913	6	1	15.486	0.088	0.071	0	0.400	1.492	31.3	1.338	4.716	120	0	0	2.60	0
126	-36.84	67.13	123.262	173.77	578.087	383.761	34.889	0.441	0.164	2.568	272	3	0	11.778	0.250	0.087	0	0.574	1.420	23.3	0.889	2.519	35	0	0	2.34	0
127[a]	-42.74	-0.92	72.393	146.20	508.311	315.706	51.534	1.138	1.225	0.842	40	0	1	2.667	0.707	0.097	0	0.387	1.425	21.3	0.822	0.810	4	0	0	1.63	0
128	-93.62	-0.35	90.428	201.36	586.184	375.543	48.427	3.574	0.039	1.817	236	0	0	20	0.144	0.074	).195	0.731	1.441	25.5	0.927	0.930	26	1	0	2.18	9.23
129	-114.26	8.07	113.358	212.63	608.120	399.028	42.552	3.467	0.028	2.346	374	1	1	22	0.102	0.084	).166	0.750	1.447	26.2	0.915	2.400	43	1	0	1.83	9.23
130[a]	-176.18	33.33	182.149	246.44	664.456	464.057	30.156	3.107	0.010	3.933	1144	4	0	28	0.036	0.102	).116	0.782	1.455	28.5	0.894	2.668	140	1	0	1.91	9.23
131	-58.96	113.55	167.719	241.36	681.556	470.273	31.955	3.707	0.007	3.449	1144	3	0	30	0.036	0.089	).117	0.794	1.479	29.1	0.919	2.349	140	1	0	1.91	9.23
132	123.47	256.52	124.102	338.90	666.487	441.819	45.902	3.837	0.017	0.686	520	0	0	26	0.083	0.075	0	0.666	1.503	38.9	0.997	0.753	61	2	0	2.23	24.7
133[a]	59.25	254.10	173.685	386.48	696.374	482.568	34.684	3.752	0.013	1.584	926	0	0	30	0.056	0.089	0	0.706	1.503	36.3	0.966	1.864	111	2	0	2.46	24.7
134[a]	81.51	197.52	125.797	266.11	652.541	431.753	41.197	3.171	0.018	1.673	560	1	0	24	0.072	0.085	0	0.785	1.498	33.6	0.927	1.859	64	1	0	2.13	12.4
135	-37.72	85.16	91.774	155.54	457.959	292.005	43.974	2.469	0.447	0.018	90	0	1	8	0.577	0.153	0	0.610	1.378	18.6	0.692	0.270	9	1	0	2.32	3.24
136[a]	-259.49	-76.99	162.386	263.43	621.800	424.687	32.430	2.433	0.029	1.743	696	1	1	24	0.083	0.112	).127	0.761	1.438	26.8	0.897	1.803	84	1	0	2.18	17.1
137	-219.95	-1.30	200.399	247.37	663.196	464.132	27.498	2.280	0.010	3.082	1314	2	0	30	0.034	0.106	).105	0.789	1.474	27.8	0.900	2.307	162	1	0	2.04	9.23
138	-172.28	149.40	249.029	314.80	721.991	517.083	22.398	3.084	0.003	3.765	2186	1	1	38	0.010	0.105	).083	0.812	1.526	31.7	0.963	2.943	267	1	0	2.19	9.23

[a]*Test set*

The stepwise multiple linear regression procedure based on the forward-selection and backward elimination methods (including the critical probability: *p*-value< 0.05 for all descriptors and for the model complete), were employed to determine the best regression models. These methods used the coefficients r, r^2, MSE, F and the *p*-value to select the best regression performance. Where r is the correlation coefficient; r^2 is the coefficient of determination; MSE is the mean squared error; F is the fisher *F*-statistic; *p*-value the significance level.

The VIF (Variance inflation factor) was defined[13] as $1/(1-r^2)$, where r was the multiple correlation coefficient for an independent variable against all other descriptors in the model. The models with a VIF greater than 5 were unstable and were eliminated, the models with VIF values between 1 and 4 may be accepted (Table 3).

Table 3. Multicollinearity statistics

	CT	PC	C%	NHD
VIF	1.042	1.540	1.424	1.171

The relationship obtained using this method corresponds to the linear combination of these descriptors (Table 3): critical temperature (CT), partition coefficient (PC), percent ratios of carbon (C%) and Number of H-Bond donors (NHD). The QSRR models built using stepwise multiple linear regression methods are represented by the following equation:

$$\text{Log(LRI)}=2.025+1.611 \times 10^{-3} \times \text{CT}-1.259\times10^{-2} \times \text{PC}+ 0.127 \times \text{C\%} +8.871\times10^{-2} \times \text{NHD} \quad (1)$$

$$N = 97;\ r = 0.900;\ r^2= 0.810;\ MSE = 0.003;\ F = 98.053;\ p\text{-value}< 0.0001.$$

In this equation, N is the number of compounds, r is the correlation coefficient, r^2 is the coefficient of determination, MSE is the mean squared error, F is the Fisher's criterion and p-value is the significance level.

It is observed that the coefficient of correlation r is high and the MSE is low, which makes it's possible to indicate that the model is more reliable. A *p*-value much smaller than 0.05 indicates that the regression equation is statistically significant, we can conclude, with confidence, that the model provides a significant amount of information[11].

The predicted log(LRI) values calculated from Eq. 1 are given in Table 4 in comparison to the observed values.

The correlation between the predicted and observed log(LRI) is shown in Figure 1. The descriptors proposed by MLR are therefore used as input parameters in the MNLR and ANN[14,15].

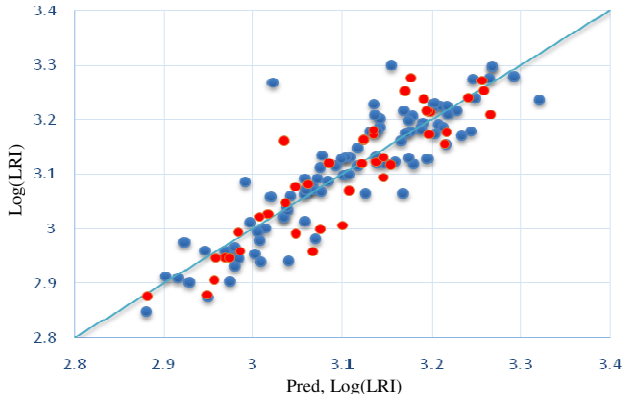


Figure 1. Graphical representation of predicted and observed linear retention indices (Log(LRI)) calculated by MLR (training set in blue; test set in red)

Multiple non-linear regression (MNLR)

We used MNLR technique to improve the QSRR by accounting for several parameters. MNLR is the most commonly used tool for the study of multidimensional data. We applied it to the data matrix constituted from the descriptors proposed by the MLR corresponding to the set of 97 molecules[16]. The resulting equation is as follows:

$$Log(LRI) = 1.277+4.240 \times 10^{-3} \times CT-2.111 \times 10^{-6} \times CT^2 -1.080\times10^{-2} \times PC - 5.185\times10^{-6} \times Exp(PC) + 5.156 \times 10^{-2} \times C\% + 6.452 \times 10^{-2} \times Log(C\%) + 8.142 \times 10^{-2} \times NHD \quad (2)$$

$N = 37$; $r = 0.908$; $r^2 = 0.824$; $MSE = 0.002$; $F = 59.452$; p-value< 0.0001.

The predicted log(LRI) values calculated from Eq. 2 are given in Table 4 in comparison to the observed values. The correlation between the predicted and observed log(LRI) is shown in Figure 2.

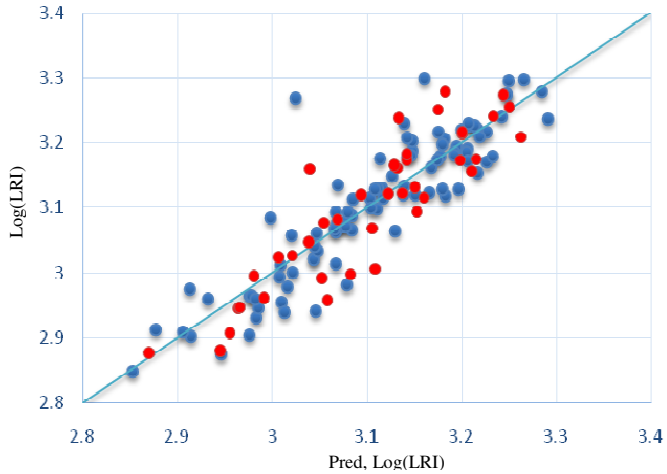


Figure 2. Predicted and observed linear retention indices (log(LRI)) using MNLR (training set in blue; test set in red)

Artificial neural networks (ANN)

In order to increase the probability of a good characterization of the molecules studied, the ANN can generate a predictive model of the QSRR relationship between the descriptors obtained from the MLR and the observed linear retention indices.

The ANN model has aroused great interests as its universal function approximators are capable of mapping any linear or nonlinear functions. The multi-layer perceptron (MLP) neuronal network model is a supervised neural network based on the original simple perceptron model with back propagation for training the network. It commonly consists of an input layer of source nodes, an output layer and one or more hidden layers of computation nodes (neurons) that increasing the learning power of the MLP model. The number of hidden neurons determines the learning capacity of MLP network. It is most recommended to select the network which performs best with the least possible number of hidden neurons[17].

The MLP model was developed using the properties of several molecules studied (Figure 3). The correlation between the predicted and observed log(LRI) is shown Figure 4. The predicted log(LRI) values calculated by ANN method are given in Table 4 to comparison to the observed values.

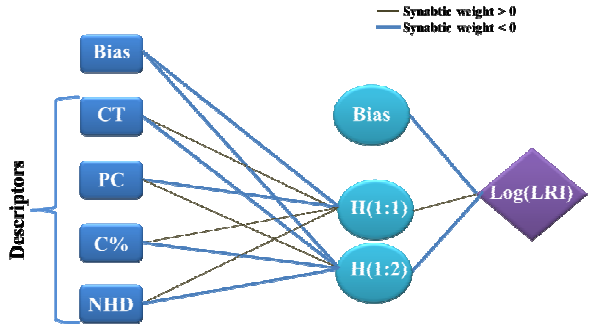


Figure 3. The architecture of the ANN method used (four input variables, two neurons in the hidden layer and one neuron to the output layer)

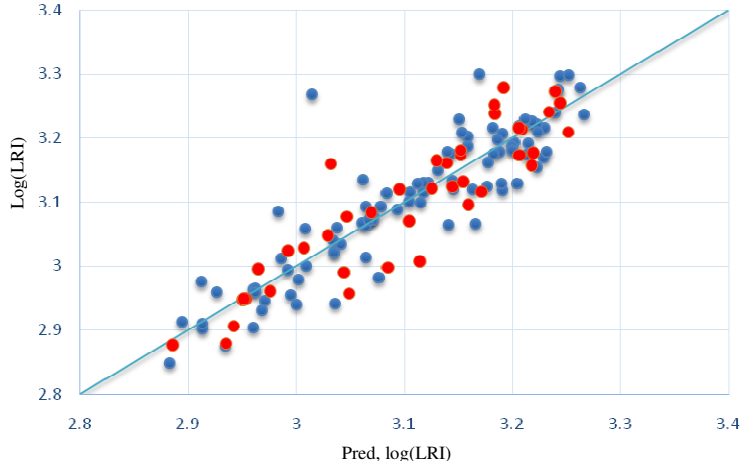


Figure 4. Observed and predicted linear retention indices using ANN (training set in blue; test set in red)

Table 4. Comparison of the observed values with those calculated by MLR, MNLR and ANN methods.

N°	Log(LRI) (obs.)	Log(LRI) (calc.)			N°	Log(LRI) (obs.)	Log(LRI) (calc.)			N°	Log(LRI) (obs.)	Log(LRI) (calc.)		
		MLR	MNLR	ANN			MLR	MNLR	ANN			MLR	MNLR	ANN
1	2.904	2.974	2.976	2.960	47	3.208	3.136	3.143	3.153	93	3.194	3.191	3.194	3.201
2	2.954	3.003	3.010	2.995	48	3.131	3.104	3.112	3.117	94	3.068	3.076	3.067	3.061
3	3.175	3.136	3.114	3.143	49[a]	3.120	3.122	3.123	3.125	95	3.120	3.139	3.138	3.145
4	3.066	3.168	3.084	3.166	50	3.115	3.117	3.117	3.117	96[a]	3.123	3.139	3.138	3.145
5	2.875	2.949	2.946	2.935	51	3.128	3.196	3.196	3.204	97	3.162	3.166	3.169	3.178
6[a]	2.879	2.949	2.946	2.935	52[a]	3.156	3.214	3.212	3.217	98	3.210	3.218	3.219	3.223
7	2.931	2.980	2.984	2.968	53	3.170	3.233	3.227	3.229	99	3.187	3.213	3.205	3.200
8	2.979	3.008	3.016	3.002	54	3.216	3.229	3.227	3.229	100[a]	3.173	3.197	3.198	3.207
9	3.023	3.034	3.044	3.035	55	3.222	3.216	3.217	3.222	101	3.179	3.177	3.178	3.189
10	3.021	3.034	3.044	3.035	56	3.226	3.210	3.208	3.216	102	3.154	3.216	3.217	3.222
11	3.063	3.059	3.067	3.065	57	3.240	3.249	3.242	3.240	103	3.134	3.077	3.070	3.062
12	3.092	3.059	3.067	3.065	58[a]	3.254	3.260	3.250	3.244	104[a]	3.251	3.171	3.175	3.183

Contd…

13	3.014	3.059	3.067	3.065	59	3.237	3.321	3.292	3.266	105	3.175	3.171	3.175	3.183
14	2.975	2.923	2.914	2.913	60	3.276	3.266	3.248	3.242	106	3.173	3.203	3.206	3.212
15[a]	2.906	2.956	2.955	2.942	61	3.296	3.269	3.250	3.244	107	3.187	3.189	3.192	3.200
16[a]	2.960	2.986	2.991	2.976	62	2.848	2.880	2.853	2.883	108	3.181	3.189	3.192	3.200
17	3.000	3.013	3.022	3.009	63	2.902	2.928	2.914	2.913	109[a]	3.069	3.108	3.106	3.105
18	3.034	3.039	3.048	3.041	64[a]	2.946	2.970	2.965	2.950	110[a]	3.209	3.267	3.264	3.252
19[a]	2.998	3.076	3.082	3.085	65	2.994	3.006	3.008	2.992	111[a]	3.214	3.198	3.200	3.208
20	3.076	3.062	3.069	3.069	66	3.040	3.038	3.045	3.034	112	3.279	3.293	3.285	3.262
21	3.077	3.062	3.069	3.069	67	3.078	3.067	3.077	3.072	113[a]	3.164	3.125	3.130	3.129
22[a]	3.082	3.062	3.069	3.069	68	3.115	3.093	3.104	3.105	114[a]	3.215	3.195	3.201	3.205
23	3.088	3.084	3.085	3.094	69	3.149	3.117	3.127	3.131	115	3.274	3.247	3.248	3.241
24	3.085	2.991	2.998	2.983	70	3.180	3.139	3.147	3.153	116	3.298	3.269	3.267	3.252
25	2.940	3.009	3.014	3.000	71[a]	3.023	3.007	3.008	2.993	117	3.216	3.170	3.176	3.182
26[a]	3.237	3.191	3.134	3.185	72	3.060	3.042	3.047	3.038	118	3.198	3.175	3.180	3.187
27	2.946	2.984	2.986	2.971	73	3.092	3.072	3.080	3.078	119	3.230	3.203	3.208	3.212
28	3.268	3.022	3.025	3.014	74	3.128	3.100	3.109	3.112	120[a]	3.131	3.145	3.151	3.155
29	3.229	3.135	3.140	3.151	75	3.099	3.102	3.111	3.115	121	2.910	2.917	2.906	2.913
30[a]	3.160	3.035	3.041	3.032	76[a]	3.161	3.124	3.133	3.140	122[a]	2.947	2.959	2.965	2.951
31	3.065	3.126	3.130	3.141	77[a]	3.173	3.135	3.142	3.152	123	2.960	2.969	2.982	2.962
32	3.120	3.148	3.151	3.163	78[a]	3.181	3.135	3.142	3.152	124	3.072	3.059	3.078	3.067
33	2.913	2.902	2.877	2.895	79	3.207	3.179	3.182	3.191	125	3.178	3.131	3.142	3.140
34	2.960	2.946	2.933	2.926	80	2.964	2.978	2.976	2.960	126	3.012	2.997	3.008	2.986
35[a]	2.994	2.984	2.980	2.965	81	2.966	2.980	2.978	2.962	127[a]	2.876	2.882	2.871	2.886
36[a]	3.027	3.018	3.021	3.007	82[a]	2.947	2.973	2.966	2.953	128	2.942	3.039	3.046	3.036
37[a]	3.077	3.048	3.056	3.047	83[a]	3.046	3.035	3.039	3.029	129	2.981	3.070	3.080	3.076
38	3.113	3.076	3.086	3.084	84	3.072	3.067	3.074	3.071	130[a]	3.095	3.145	3.153	3.159
39	3.058	3.020	3.021	3.009	85	3.191	3.207	3.206	3.215	131	3.118	3.180	3.183	3.191
40[a]	3.119	3.086	3.095	3.096	86[a]	3.241	3.242	3.233	3.234	132	3.129	3.175	3.180	3.190
41	3.131	3.108	3.116	3.122	87[a]	3.272	3.257	3.244	3.240	133[a]	3.176	3.217	3.215	3.219
42[a]	2.990	3.048	3.053	3.045	88	3.220	3.201	3.200	3.206	134[a]	3.116	3.155	3.160	3.172
43	3.124	3.159	3.166	3.176	89	3.226	3.218	3.214	3.218	135	2.796	2.840	2.793	2.867
44	3.187	3.142	3.148	3.159	90	3.100	3.107	3.105	3.104	136[a]	3.006	3.102	3.110	3.115
45	3.202	3.142	3.148	3.159	91[a]	2.957	3.068	3.058	3.049	137	3.300	3.155	3.161	3.170
46[a]	3.277	3.178	3.182	3.193	92	3.133	3.138	3.139	3.145	138	3.178	3.244	3.233	3.232

[a]*Test Set*

Internal validation

Cross-Validation

The cross-validation statistical procedure can be used to evaluate the predictive power of QSRR models. The leave-one-out (LOO) procedure successively removes one molecule from the training set containing n molecules. A QSRR model is constructed on a "n-1" set of compounds and the molecule removed is predicted by the model. This procedure is repeated "n" times in order to predict the retention of all molecules.

The QSRR models expressed by the equations of MLR and MNLR are validated by its appreciable values of r_{CV}^2 (Table 5) obtained using the LOO procedure. The value of r_{CV}^2 greater than 0.5 is the basic condition for qualifying a QSRR model as valid.

Table 5. r_{CV}^2 values obtained by the LOO

	MLR	MNLR
r_{CV}	0.890	0.883
r_{CV}^2	0.792	0.780

We use cross validation as an internal test of the quality of MLR and MNLR models. The performance of models was good and was characterized by r_{CV}^2 values; 0.792 for the MLR and 0.780 for MNLR method (Table 5).

y-Randomization test

To ensure that the developed models are robust and not derive due to chance, the *y*-randomization test was performed on the training set data as recommended[18]. In this test, models are generated by randomly scrambling the dependent variable (Log(LRI)) while keeping the independent variable (descriptors) unchanged. The resulting models are expected to have significantly low r^2 and cross validated r_{CV}^2 values for several trials, which confirm that the developed models are robust. We performed 100-y-randomization tests and observed that for all the models, the values of r^2 and r_{CV}^2 were <0.5 (Figure 5). This test confirms that the developed models are robust and not derived merely due to chance[19].

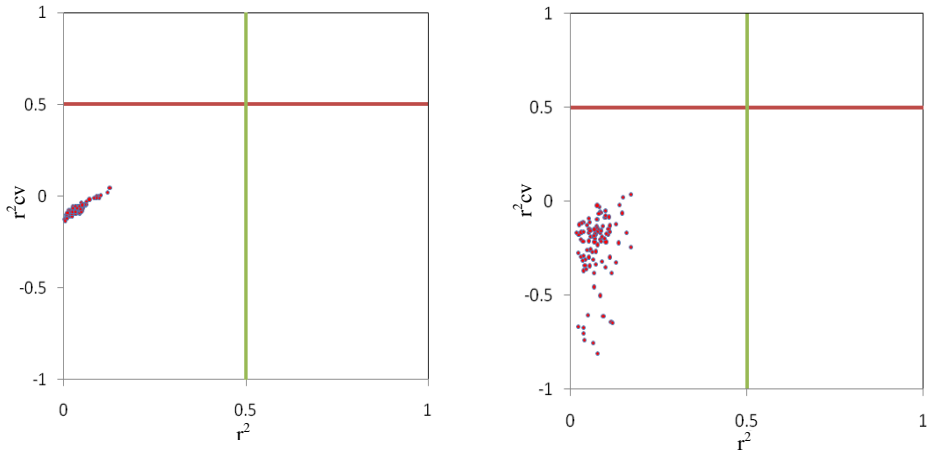


Figure 5. *y*-Randomization plot of MLR and MNLR models

External validation

To estimate the predictive power of the MLR, MNLR and ANN models, we have used a set of compounds that have not been used in the training set to establish the models. The equations of the models established are used to predict the linear retention indices (Log(LRI)) of the remaining (41 molecules). The main performance parameters for the three models are shown in Table 6.

Table 6. Comparison of MLR, MNLR and ANN models

		MLR	MNLR	ANN
Training Set	*r*	0.900	0.908	0.903
	r^2	0.810	0.824	0.816
	MSE	0.003	0.002	0.002
Test Set	*r*	0.903	0.899	0.909
	r^2	0.816	0.804	0.826
	MSE	0.002	0.003	0.002

The results obtained by the three models, are very sufficient to conclude the performance of models; it's confirmed by the test done with the 41 compounds.

A comparison of the quality of MLR, MNLR and ANN models shows that the three approaches have better predictive capability gives better results. MLR, MNLR and ANN were able to establish a satisfactory relationship between the molecular descriptors and the linear retention indices (Log(LRI)) of the studied compounds.

Domain of applicability

Evaluation of the applicability domain of the QSRR model is considered as an important step to establish that the model is reliable to make predictions within the chemical space for which it was developed[20]. There are several methods for defining the applicability domain of a QSRR model, but we used the most commonly used leverage approach in this study[21]. Leverage of a given chemical compound hi is defined as:

$$h_i = x_i^T(X^T X)^{-1}x_i \,(i = 1. . . .n)$$

Where x_i is the descriptor row of the query compound and X is the descriptor matrix of the training set compounds used to develop the model. As a prediction tool, the warning leverage h* is defined as:

$$h^* = 3(p + 1)/n$$

Where n is the number of training compounds and p is the number of descriptors in the model. The test compounds with leverages hi<h* are considered to be reliably predicted by the model. The Williams plot is used to interpret the applicability domain of the model. The domain of reliable prediction for external test set compounds is defined as compounds which have leverage values within the threshold (hi<h*) and standardized residuals no greater than 3 units (±δ). Test set compounds where (hi>h*) are considered to be unreliably predicted by the model due to substantial extrapolation. For the training set, the Williams plot is used to identify compounds with the greatest structural influence (hi>h*) in developing the model.

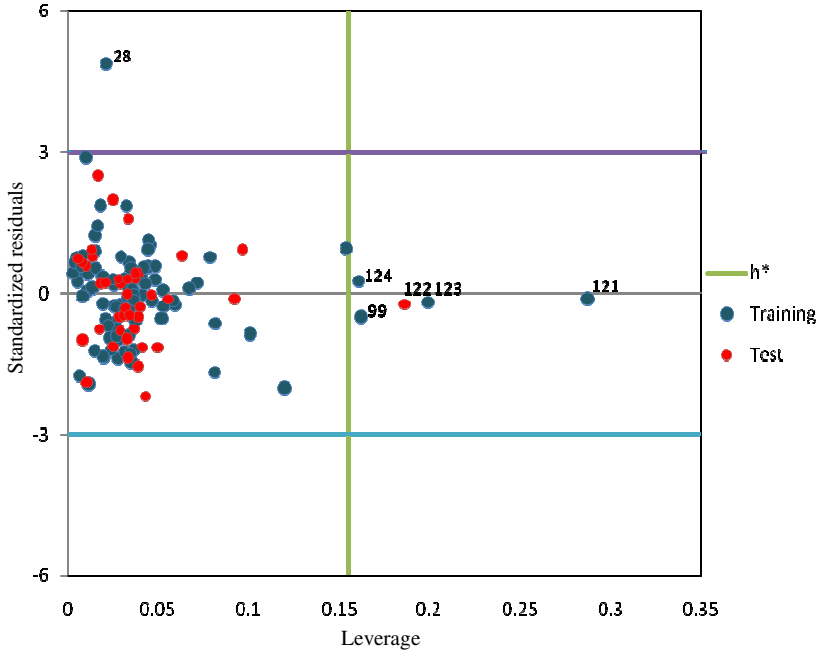


Figure 6. Williams plot to evaluate the applicability domain of MLR model

From the Williams plot (Figure 6), it is obvious that all the compounds in the dataset are within the applicability domain of the model (the warning leverage limit is 0.155) except one test set compound (122) and four training compounds (99, 121, 123 and 124), these four compounds have their leverage values greater than the warning h* value and could be high leverage compounds influencing the performance of the model. However, their standard residual values are very low and within the established limit[17]. As a result, these four compounds could be considered as influential in fitting the model performance but not necessarily outliers to be deleted from the training dataset and thus the model can be applied with confidence within the defined applicability domain.

For all the compounds in the training and test sets, their standardized residuals are smaller than three standard deviation units (3±δ), except one in the training set (compound N°28). Therefore, the predicted linear retention indices (Log(LRI)) by the developed MLR model is reliable

Proposed novel compounds

QSRR correlates retention data with the physicochemical properties of a group of compounds. It has been frequently used to predict proprieties of new compounds and to design compounds with desired properties.

The developed Eq. 1 can be used for the designing of new volatile compounds derivatives with improved linear retention indices(Log(LRI)).

Comparing *t*-test and standardized coefficient values of descriptors (Table 7) indicates that the influences of the Critical temperature (CT), the Partition coefficient (PC) and the number of *H*-Bond donors (NHD) on log(LRI) are stronger than this of the Percent ratios of carbon (C%).

Table 7. *T-test* and standardized coefficient values of descriptors for equation (1)

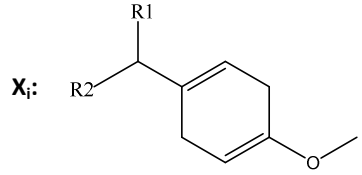
	Standardized coefficient	*t*-test	Sign.
CT	0.000	17.620	**< 0.0001**
PC	0.003	-4.534	**< 0.0001**
C%	0.044	2.905	0.005
NHD	0.012	7.190	**< 0.0001**

The equation (1) of the MLR method indicated the positive correlation of the critical temperature (CT) and the number of H-Bond donors (NHD), it indicated also the negative correlation of the partition coefficient (PC).

The obtained results show that, to increase linear retention indices of volatile compounds, we will increase critical temperature (CT) and the number of *H*-Bond donors (NHD) and we will decrease the partition coefficient (PC). Moreover, to decrease linear retention indices, we will decrease critical temperature (CT) and the number of *H*-Bond donors (NHD) and we will increase the partition coefficient (PC) of this molecules, by adding suitable substituent's and calculated their linear retention indices (Log(LRI)) using the Eq. 1.

The structures of the designed compounds and their parameter values calculated by the same methods as well as log(LRI) values theoretically predicted by the MLR model (Eq. 1) are listed in Table 8.

Table 8. Values of descriptors, linear retention indices (Log(LRI)) and leverages (h) for the new designed compounds



	Designed compounds	CT	PC	C%	NHD	Log(LRI)	Leverage, h
X_1	R1=OH; R2=CH_3	692.56	0.676	0.701	1	3.310	0.050
X_2	R1=OH; R2=OH	719.10	-0.462	0.615	2	3.445	0.163
X_3	R1=OH; R2=C_2H_5	706.75	1.205	0.714	1	3.328	0.054
X_4	R1=OH; R2=$CH(CH_3)_2$	710.14	1.604	0.725	1	3.330	0.055
X_5	R1=OH; R2=C_3H_7	719.95	1.734	0.725	1	3.344	0.060
X_6	R1=CH_2OH; R2=CH_3	714.14	1.095	0.714	1	3.341	0.059
X_7	R1=CH_2OH; R_2=CH_2OH	754.84	-0.236	0.562	2	3.493	0.183
X_8	R1=CH_2OH;R2=C_2H_5	726.86	1.624	0.725	1	3.356	0.064
X_9	R1=CH_2OH; R2=$CH(CH_3)_2$	729.84	2.023	0.734	1	3.357	0.066
X_{10}	R1=CH_2OH; R2=$CH(CH_3OH)$	752.14	0.073	0.666	2	3.498	0.181
X_{11}	R1=CH_2OH; R2=C_3H_7	738.77	2.153	0.734	1	3.370	0.071
X_{12}	R1=CH_2OH; R2=Cl	740.90	0.759	0.573	1	3.371	0.081
X_{13}	R1=CH_2OH; R2=F	694.69	0.419	0.628	1	3.307	0.051
X_{14}	R1=CH_2OH; R2=Br	762.31	0.899	0.464	1	3.389	0.116
X_{15}	R1=OH; R2=Br	745.15	1.120	0.439	1	3.356	0.111
X_{16}	R1=OH; R2=C_6H_5	799.07	2.015	0.778	1	3.475	0.121
X_{17}	R1=CH_2OH; R2=C_6H_5	810.61	2.134	0.782	1	3.492	0.131

From the predicted properties, it has been observed that the designed compounds X_1, X_2, X_3, X_4, X_5, X_6, X_7,X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, $X_{15,}$ X_{16} and X_{17} have higher Log(LRI) values than the existing compounds in the case of the 138 studied compounds (Table 1).

The leverage values (h) calculated by Eq. 1 of the MLR for the new designed compounds are displayed in Table 8. Only three compounds X_2, X_7 and X_{10} are defined as outliers and consequently they are not be considered, because they have higher leverage which is greater than h* (h*=0.155)[11].

Conclusion

In this study, multiple linear and non-linear regression and artificial neural networks were used to construct quantitative structure retention relation models of volatile compounds for their linear retention indices (LRI). MLR analysis produced more predictive, informative and significantly improved QSRR model. All QSRR models provide a reasonably good correlation coefficient (r≥0.900). The validation and predictive ability of the models were examined by the leave one-out cross-validation, *Y*-randomization and external validation. The three methods indicated that the resulting multiparametric QSRR models possess high prediction ability and low over fitting. The applicability domain of the MLR model was defined.

We conclude that the most important finding about this research is that we have been able to design and proposed some new compounds with high linear retention indices (LRI) values than the existing ones by adding suitable substituent's and calculated their linear retention indices using regression equation.

Acknowledgment

We are grateful to the "Association Marocaine des ChimistesThéoriciens" (AMCT) for its pertinent help concerning the programs.

References

1. Ghavami R and Faham S, *Chromatographia*, 2010, **72(9-10),** 893-903; DOI:10.1365/s10337-010-1741-4
2. **Dai J,** Jin L, Yao S H and Wang L, *Chemosphere,* 2001, **42(8),** 899-907. DOI:10.1016/S0045-6535(00)00181-8
3. **Qin L,** Liu Sh Liu H and Tong C, *J Chromatogr A*, 2009, **1216(27),** 5302–5312. DOI:10.1016/j.chroma.2009.05.016
4. Kaliszan R, Recent advances in quantitative structure retention relationships; separation methods in drug synthesis and purification. In: Valko K, (Ed), Elsevier, Amsterdam, 2000.
5. Hu R J, Liu H X, Zhang R S, Xue C X, Yao X J, Liu M C, Hu Z D and Fan B T, *Talanta*, 2005, **68(1),** 31-39; DOI:10.1016/j.talanta.2005.04.034
6. **López-Pérez O,** Picon A and Nuñez M, *Food Res Int.*, in press; 2016, 99(3), 1002-1010; DOI:10.1016/j.foodres.2016.12.013
7. **SPSS 19.0**, http://www.ibm.com/analytics/fr/fr/technology/spss/.
8. A**CDLABS 10**, "Advanced Chemistry Development, Inc., Toronto, ON, Canada," **2015**, http://www.acdlabs.com
9. Marvin Sketch 5.11.4. "Chem Axon." **2012**, http://www.chemaxon.com
10. ChemBioOffice, *PerkinElmer Informatics*, **2010**, http://www.cambridgesoft.com
11. Chtita S, Ghamali M, Hmamouchi R, Elidrissi B, Bourass M, Larif M, Bouachrine M and Lakhlifi T, *Adv Phys Chem*., 2016, 1-16. DOI:10.1155/2016/5137289
12. Gramatica P, *Comb Sci.*, 2007, **26(5),** 694-701; DOI:10.1002/qsar.200610151
13. Belhassan A, Chtita S, Lakhlifi T and Bouachrine M, *Electro J Chem*., 2017**, 9(4),** 234-237; DOI:10.17807/orbital.v9i4.978
14. Chtita S, Hmamouchi R, Larif M, Ghamali M, Bouachrine M and Lakhlifi T, *J Taibah Univ Sci.*, 2016, **10(6),** 868-876; DOI:10.1016/j.jtusci.2015.04.007.
15. Belhassan A, Chtita S, Lakhlifi T and Bouachrine M, *J Taibah Univ Sci*., 2017, **11(6),** 1030-1046; DOI:10.1016/j.jtusci.2017.05.004.
16. Chtita S, Larif M, Ghamali M, Bouachrine M and Lakhlifi T, *J Taibah Univ Sci*., 2015, **9(2)**, 143-154; DOI:10.1016/j.jtusci.2014.10.006.
17. Bastani D, Hamzehie M E, Davardoost F, Mazinani S and Poorbashiri A, *Fluid Phase Equilibria*, 2013, **354,** 6-11; DOI:10.1016/j.fluid.2013.05.017
18. Roy K and Mitra I, *Comb Chem High Throughput Screen*, 2011, **14(6),** 450-474; DOI:10.2174/138620711795767893
19. Garg R and Smith C J, *Food Chem Toxicol.*, 2014, **69,** 252-259; DOI:10.1016/j.fct.2014.03.035
20. Eriksson L, Jaworska J, Worth A P, Cronin M T D, McDowell R M and Gramatica P, *Environ Health Perspect*, 2003, **111(10)**, 1361-1375; http://refhub.elsevier.com/S02-78-6915(14)00164-1/h0065
21. Gramatica P, *Comb Sci*., 2007, **26,** 694-701; http://refhub.elsevier.com/S0278-6915(14)00164-1/h0085