Global Journal of Molecular Sciences 2 (1): 12-17, 2007
ISSN 1990-9241

Optimal Descriptors Based on Extended Connectivity and Codes of Cyles: QSPR of Hydrocarbon Normal Boiling Points

[1]E.A. Castro, [2]A.A. Toropov, [2]A.P. Toropova and [2]R.Z. Akhmerov

[1]INIFTA, Theoretical Chemistry Division, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina
[2]Uzbekistan Academy of Science, Institute of Geology and Geophysics,
Khodzhibaev street 49, 100041 Tashkent, Uzbekistan

Abstract: (0, 1)-codes reflected presence / absence of six- and five-members cycles produce considerable improving statistical characteristics of quantitative structure–property relationships between optimal descriptors and normal boiling points of alkanes, alkylbenzenes and polyaromatic hydrocarbons.

Key words: QSPR · normal boiling points · hydrocarbons · cycle code · optimal descriptors

INTRODUCTION

Our real world is rather uncertain. Suppose we are carrying out a chemical reaction, which allows us to obtain a product. In the very beginning we observe a complete uncertainty regarding the molecule. In fact, we have no information about its composition, the constitution of the molecular skeleton, its stereochemical features, its physical properties and biological activities and so on. Gradually, performing routine suitable experiments, we gather data. Then the acquisition of the structural information is complete there is no uncertainty, at least about its structure. The information about a real physical system is a measure of decreasing uncertainty of the system by means of some physical acitivities.

We can define knowledge as the perception of the logical relations among the structures of the information. Any systematic treatment of information needs some previous knowledge. Therefore, research is always an iterative process, as depicted in Scheme 1 [1].

The central problem in theory of Quantitative Structure Property-Activity Relationships (QSPR/QSAR) is to convert chemical structures in to mathematical molecular descriptors that are relevant to the physico-chemical property or to the mechanism of the biological activity. Topological indices are among the best descriptors, although there are other sort of them, such as Molecular Orbital indices and other properties/activities which are also employed in these studies. It is a well known fact that molecular structure is one of the basic concepts of chemistry since properties and molecular behavior follow from their structures. In particular, properties of a molecule are a consequence of a complicated interplay of its topology (i.e. atomic connectivity), metric characteristics (bond lengths, valence and torsional angles) and detailed dynamics of electrons and nuclei. Finding out how various molecular features depend on molecular structure is one of the central fields of chemistry and particularly the main subject of QSAR/QSPR studies [2].

The topological indices are those structure molecular descriptors derived from a graph theoretical representation of molecules. These molecular descriptors should carry out most of the desired attributes for topological indices. Flexible topological descriptors make up a quite efficient set of variables to employ in QSAR/QSPR studies which have been employed successfully in many cases to predict physico-chemical properties and biological activities [3-10].

In a recent study [11] on the comparison of QSPR models based on hydrogen-filled graphs and on graphs of atomic orbitals, optimal descriptors have been calculated without taking into account the presence/absence of cycles in the molecular structure. The aim of the present study is to estimate the efficacy of codes of six- and five-member cycles in constructing optimal descriptors in order to get a better molecular description. We have chosen a set of 140 hydrocarbons to predict their normal boiling points.

Correspondoing Author: Dr. E.A. Castro, INIFTA, Theoretical Chemistry Division, Suc.4, C.C. 16, La Plata 1900, Buenos Aires, Argentina

METHODS

Descriptors used in the present study have been calculated as

$$^{1}DCW(a_k,EC_k)=\{\sum_{k=1}^{n} CW(a_k)+\sum_{k=1}^{n} CW(^{x}EC_k)\}^{m} \quad (1)$$

$$^{2}DCW(a_k,EC_k)=\{CW(CC)+\sum_{k=1}^{n} CW(a_k)+\sum_{k=1}^{n} CW(^{x}EC_k)\}^{m} \quad (2)$$

where:

a_k is chemical element (C or H) in hydrogen-filled molecular graph (HFG) that is presented by k-th vertex in the graph,

$^{x}EC_k$ is the extended connectivity of x-th order (x = 0,1,2),

CC is the code of cycles and they were calculated as shown in Table 1,

$CW(a_k)$ is the correlation weight of presence a_k in HFG,

$CW(^{x}EC_k)$ is the correlation weight for a given extended connectivity value,

CW(CC) is the correlation weight of the code of cycles,

n is the number of vertex in the HFG and

m = 0.5.

Numerical data on the $CW(a_k)$, $CW(^{x}EC_k)$ and CW(CC) have been calculated via Monte Carlo optimization method, i.e., we look for those values of the CWs producing maximal correlation coefficient between ^{q}DCW and normal boiling points (NBPs) of hydrocarbon of the training set have been obtained by the corresponding optimization procedure. From the data one can then calculate the desired physical chemistry property by the Least Square method NBP model

$$NBP = C_0 + C_1\ ^{q}DCW(a_k,EC_k) \quad (3)$$

Table 1: Definition of the (0,1) cycle codes (CC)

Situation in molecular structure	Numerical value of the CC
There is no cycle	C00
Six member cycle (one or more)	C10
Five member cycle (one or more)	C01
Both six-member and five-member cycles	C11

Predictive potential of the model must be validated with an external test set. Recently these hydrocarbons have been examined in Ref. 11. We have resorted to the splitting of the whole molecular set into a training and a test set from the study. Choice of m = 0.5 is based on fact that often correlation between normal boiling points and descriptors are non linear [12,13]. We have tested m = 1 and m = 0.5. Statistical characteristics in case of m = 0.5 were better than those corresponding to m = 1. Details on models with m = 1 will not be examined further in this paper. Calculation of the extended connectivity of increasing orders has been described in Ref. 14. Version of the Cycle codes (CC) are shown in Table 1 and it is a particular case of the generalized CC definition used in [15].

RESULTS AND DISCUSSION

From Table 2 one can see that ^{2}DCW modeling gives a model of normal boiling points of better statistical

Table 2: Statistical characteristics on the ^{1}DCW- and ^{2}DCW-modeling

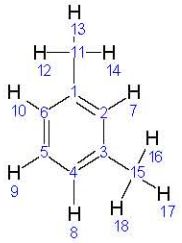
Probe		N_{op}*	C_1	C_0	r^2	s, °C	F	r^2	s, °C	F
^{1}DCW-modeling										
1	°EC	5	50.759	-378.100	0.9917	15.526	8140	0.9935	12.595	10351
2			56.014	-378.098	0.9917	15.526	8140	0.9935	12.596	10351
3			53.229	-378.123	0.9917	15.526	8140	0.9935	12.595	10353
1	^{1}EC	10	79.442	-364.654	0.9958	11.065	16092	0.9958	9.874	16260
2			86.660	-364.808	0.9958	11.065	16092	0.9958	9.874	16262
3			81.802	-364.789	0.9958	11.065	16092	0.9958	9.877	16252
1	^{2}EC	23	113.195	-497.071	0.9991	5.207	72900	0.9959	9.735	16488
2			118.798	-497.180	0.9991	5.208	72893	0.9959	9.742	16467
3			116.924	-495.901	0.9991	5.210	72830	0.9959	9.690	16642
^{2}DCW-modeling										
1	°EC	8	153.759	-656.597	0.9979	7.818	32303	0.9980	6.893	33635
2			149.916	-635.847	0.9979	7.834	32177	0.9980	6.911	33477
3			153.089	-661.713	0.9979	7.815	32331	0.9980	6.896	33603
1	^{1}EC	13	117.526	-448.805	0.9988	5.857	57619	0.9990	4.821	68071
2			119.400	-442.533	0.9988	5.866	57434	0.9990	4.830	67744
3			110.481	-438.225	0.9988	5.871	57332	0.9990	4.825	67868
1	^{2}EC	26	115.581	-537.107	0.9992	4.725	88550	0.9970	8.277	22824
2			108.701	-534.546	0.9992	4.735	88180	0.9973	7.965	24661
3			113.946	-534.428	0.9992	4.724	88599	0.9972	8.107	23802

*) N_{op} is number of optimized parameters.

Table 3: Correlation weights for calculation $^2DCW(a_k,{}^1EC_k)$

HFG invaraint	CWs on probe 1	CWs on probe 2	CWs on probe 3
Chemical elements, a_k			
H	0.03124	0.17661	0.02428
C	0.00000	0.00000	0.00000
Extended connectivity of first order, ^1EC_k			
0003	2.46163	2.24314	2.78192
0004	0.26443	0.12287	0.32249
0006	2.45652	2.36583	2.74739
0007	0.02004	0.00309	0.00068
0009	3.66781	3.52048	4.09422
0010	1.75035	1.64111	1.89861
0013	3.03027	2.84582	3.30193
0016	4.32879	4.07044	4.71470
(0,1) cycle codes, CC			
C00	7.83707	7.22255	8.19960
C10	4.86283	4.29464	4.78143
C11	2.84091	2.35837	2.51194

Table 4: Calculation of the $^2DCW(a_k,{}^1EC_k)$ for m-xylene



CW(CC) = CW(C10) = 4.86283

$$CW(CC)+\sum_{k=1}^{n} CW(a_k)+\sum_{k=1}^{n} CW(^{\pi}EC_k)=25.10227$$

$25.20227^{0.5} = 5.01022$

No.	a_k	^1EC_k	$CW(a_k)$	$CW(^1EC_k)$
1	C	10	0.00000	1.75035
2	C	7	0.00000	0.02004
3	C	10	0.00000	1.75035
4	C	7	0.00000	0.02004
5	C	7	0.00000	0.02004
6	C	7	0.00000	0.02004
7	H	3	0.03124	2.46163
8	H	3	0.03124	2.46163
9	H	3	0.03124	2.46163
10	H	3	0.03124	2.46163
11	C	6	0.00000	2.45652
12	H	4	0.03124	0.26443
13	H	4	0.03124	0.26443
14	H	4	0.03124	0.26443
15	C	6	0.00000	2.45652
16	H	4	0.03124	0.26443
17	H	4	0.03124	0.26443
18	H	4	0.03124	0.26443

Table 5: Experimental and calculated with Eq. (4) values of hydrocarbon normal boiling points on training and test sets

No.	Structures	DCW	NBP_{expr}	NBP_{calc}	NBP_{expr} - NBP_{calc}
Training set					
1	Ethane	3.10664	-88.60000	-83.70767	-4.89233
2	Propane	3.46307	-42.10000	-41.82001	-0.27999
3	2,2-Dimethylpropane	3.97418	9.50000	18.24563	-8.74563
4	2-Methylbutane	4.02814	27.80000	24.58701	3.21299
5	2-Methylpropane	3.72615	-11.70000	-10.90285	-0.79715
6	n-Butane	3.78610	-0.50000	-3.85753	3.35753
7	3-Methylpentane	4.30901	63.30000	57.59486	5.70514
8	n-Hexane	4.36096	68.70000	63.70002	4.99998
9	3,3-Dimethylpentane	4.52521	86.10000	83.00268	3.09732
10	2,3-Dimethylpentane	4.52315	89.80000	82.76059	7.03941
11	3-Methylhexane	4.57267	91.80000	88.58018	3.21982
12	3-Ethylpentane	4.57267	93.50000	88.58018	4.91982
13	2,2,3-Trimethylpentane	4.72957	109.80000	107.01907	2.78093
14	2,3,3-Trimethylpentane	4.72957	114.80000	107.01907	7.78093
15	3,3-Dimethylhexane	4.77694	112.00000	112.58599	-0.58599
16	3-Ethyl-3-methylpentane	4.77694	118.30000	112.58599	5.71401
17	3-Ethyl-2-methylpentane	4.77500	115.60000	112.35800	3.24200
18	2,2-Dimethylbutane	4.25861	49.70000	51.67185	-1.97185
19	2,2,3,3-Tetramethylpentane	4.92734	140.30000	130.26100	10.03900
20	2,2,3,4-Tetramethylpentane	4.92545	133.00000	130.03888	2.96112
21	2,3,3-Trimethylhexane	4.97097	137.70000	135.38839	2.31161
22	2,2,4,4-Tetramethylpentane	4.92734	122.30000	130.26100	-7.96100
23	2,2,4-Trimethylhexane	4.97097	126.50000	135.38839	-8.88839
24	2,4,4-Trimethylhexane	4.97097	130.60000	135.38839	-4.78839
25	3,3-Diethylpentane	5.01606	146.20000	140.68737	5.51263
26	2,4-dimethyl-3-ethylpentane	4.96909	136.70000	135.16746	1.53254
27	3-Ethyl-4-methylhexane	5.01421	140.40000	140.46996	-0.06996
28	4-Ethyl-2-methylhexane	5.01421	133.80000	140.46996	-6.66996
29	2-Methyloctane	5.05892	143.30000	145.72428	-2.42428
30	3-Methyloctane	5.05892	144.20000	145.72428	-1.52428
31	4-Ethylheptane	5.05892	141.20000	145.72428	-4.52428
32	2,2-Dimethylheptane	5.01606	132.70000	140.68737	-7.98737
33	2,5-Dimethylheptane	5.01421	136.00000	140.46996	-4.46996
34	2,6-Dimethylheptane	5.01421	135.20000	140.46996	-5.26996
35	3,5-Dimethylheptane	5.01421	136.00000	140.46996	-4.46996
36	3-Methyl-3-ethylhexane	5.01606	140.60000	140.68737	-0.08737
37	Benzene	4.46546	80.10000	75.98086	4.11914
38	Toluene	4.74566	110.60000	108.90996	1.69004
39	m-Xylene	5.01022	139.10000	140.00105	-0.90105
40	p-Xylene	5.01022	138.40000	140.00105	-1.60105
41	1-Methyl-3-ethylbenzene	5.18892	161.30000	161.00188	0.29812
42	1-Methyl-4-ethylbenzene	5.18892	162.00000	161.00188	0.99812
43	1,3,5-Trimethylbenzene	5.26149	164.70000	169.53030	-4.83030
44	1,2,3,4-Tetramethylbenzene	5.50129	205.00000	197.71160	7.28840
45	1,3-Diethylbenzene	5.36168	181.10000	181.30463	-0.20463
46	1,4-Diethylbenzene	5.36168	183.80000	181.30463	2.49537
47	1-Methyl-4-n-propylbenzene	5.40986	183.80000	186.96675	-3.16675
48	1,2-Dimethyl-3-ethylbenzene	5.43194	193.90000	189.56159	4.33841
49	1,3-Dimethyl-4-ethylbenzene	5.43194	188.40000	189.56159	-1.16159
50	1,3-Dimethyl-5-ethylbenzene	5.43194	183.80000	189.56159	-5.76159
51	1,2,4,5-Tetramethylbezene	5.50129	196.80000	197.71160	-0.91160

Table 5: Continued

52	Naphthalene	5.68346	218.00000	219.12022	-1.12022
53	Acenaphthylene	6.13314	270.00000	271.96661	-1.96661
54	Acenaphthene	6.10516	279.00000	268.67840	10.32160
55	Fluoranthene	7.06943	383.00000	381.99941	1.00059
56	Pyrene	7.21102	393.00000	398.63907	-5.63907
57	Benzo(c)fluorene	7.24498	406.00000	402.63005	3.36995
58	Benzo(ghi)fluoranthene	7.57050	422.00000	440.88516	-18.88516
59	Benz(a)anthracene	7.55147	425.00000	438.64875	-13.64875
60	Dibenz(a,j)anthracene	8.32983	531.00000	530.12162	0.87838
61	Cyclopenta(cd)pyrene	7.54785	439.00000	438.22333	0.77667
62	Benzo(k)fluoranthene	7.89546	481.00000	479.07446	1.92554
63	Perylene	8.02248	497.00000	494.00185	2.99815
64	Anthanthrene	8.46734	547.00000	546.28180	0.71820
65	Indeno(1,2,3-cd)pyrene	8.34709	534.00000	532.15002	1.84998
66	Dibenz(a,c)anthracene	8.32983	535.00000	530.12162	4.87838
67	Picene	8.32983	519.00000	530.12162	-11.12162
68	Coronene	8.88997	590.00000	595.94927	-5.94927
69	Dibenzo(a,i)pyrene	8.75909	594.00000	580.56826	13.43174
70	Dibenzo(a,l)pyrene	8.75909	595.00000	580.56826	14.43174
Test set					
1	4-Methylheptane	4.82193	117.70000	117.87321	-0.17321
2	n-Pentane	4.08366	36.10000	31.11172	4.98828
3	Dibenzo(a,e)pyrene	8.75909	592.00000	580.56826	11.43174
4	Dibenzo(a,h)pyrene	8.75909	596.00000	580.56826	15.43174
5	Dibenz(a,h)anthracene	8.32983	535.00000	530.12162	4.87838
6	Benzo(ghi)perylene	8.46734	542.00000	546.28180	-4.28180
7	Indeno(1,2,3-cd)fluoranthene	8.34709	531.00000	532.15002	-1.15002
8	Benz(a)pyrene	8.02248	496.00000	494.00185	1.99815
9	Benzo(e)pyrene	8.02248	493.00000	494.00185	-1.00185
10	Naphthacene	7.55147	440.00000	438.64875	1.35125
11	Triphenylene	7.55147	429.00000	438.64875	-9.64875
12	Benzo(b)fluoranthene	7.89546	481.00000	479.07446	1.92554
13	Chrysene	7.55147	431.00000	438.64875	-7.64875
14	Benzo(a)fluorene	7.24498	403.00000	402.63005	0.36995
15	Benzo(b)fluorene	7.24498	398.00000	402.63005	-4.63005
16	Fluorene	6.33469	294.00000	295.65277	-1.65277
17	4H-Cyclopenta(def)phenanthrene	6.88941	359.00000	360.84346	-1.84346
18	Benzo(j)fluoranthene	7.89546	480.00000	479.07446	0.92554
19	Phenanthrene	6.68305	338.00000	336.59204	1.40796
20	Anthracene	6.68305	340.00000	336.59204	3.40796
21	1,4-Dimethyl-2-ethylbenzene	5.43194	186.90000	189.56159	-2.66159
22	1,2,3,5-Tetramethylbenzene	5.50129	198.20000	197.71160	0.48840
23	1,2-Dimethyl-4-ethylbenzene	5.43194	189.80000	189.56159	0.23841
24	1,3-Dimethyl-2-ethylbenzene	5.43194	190.00000	189.56159	0.43841
25	1-Methyl-2-n-propylbenzene	5.40986	184.80000	186.96675	-2.16675
26	1-Methyl-3-n-propylbenzene	5.40986	181.80000	186.96675	-5.16675
27	n-Butylbenzene	5.38770	183.30000	184.36250	-1.06250
28	1,2-Diethylbenzene	5.36168	183.40000	181.30463	2.09537
29	1,2,3-Trimethylbenzene	5.26149	176.10000	169.53030	6.56970
30	1,2,4-Trimethylbenzene	5.26149	169.40000	169.53030	-0.13030
31	n-Propylbenzene	5.16582	159.20000	158.28717	0.91283
32	1-Methyl-2-ethylbenzene	5.18892	165.20000	161.00188	4.19812

Table 5: Continued

33	Ethylbenzene	4.93396	136.20000	131.03898	5.16102
34	o-Xylene	5.01022	144.40000	140.00105	4.39895
35	3,3,4-Trimethylhexane	4.97097	140.50000	135.38839	5.11161
36	2,3,4-Trimethylhexane	4.96909	139.00000	135.16746	3.83254
37	3,3-Dimethylheptane	5.01606	137.30000	140.68737	-3.38737
38	3,4-Dimethylheptane	5.01421	140.60000	140.46996	0.13004
39	2,3-Dimethylheptane	5.01421	140.50000	140.46996	0.03004
40	2,4-Dimethylheptane	5.01421	133.50000	140.46996	-6.96996
41	4-Methyloctane	5.05892	142.50000	145.72428	-3.22428
42	3-Ethylheptane	5.05892	143.00000	145.72428	-2.72428
43	3,4-Dimethylhexane	4.77500	117.70000	112.35800	5.34200
44	n-Nonane	5.10324	150.80000	150.93276	-0.13276
45	2,3,5-Trimethylhexane	4.96909	131.30000	135.16746	-3.86746
46	3-ethyl-2-methyl-hexane	5.01421	138.00000	140.46996	-2.46996
47	2,2,5-Trimethylhexane	4.97097	124.10000	135.38839	-11.28839
48	4,4-Dimethylheptane	5.01606	135.20000	140.68737	-5.48737
49	2,3-dimethyl-3-ethylpentane	4.97097	142.00000	135.38839	6.61161
50	2,2,3-Trimethylhexane	4.97097	133.60000	135.38839	-1.78839
51	2,2-dimethyl-3-ethylpentane	4.97097	133.80000	135.38839	-1.58839
52	2,3,3,4-Tetramethylpentane	4.92545	141.60000	130.03888	11.56112
53	3-Ethylhexane	4.82193	118.50000	117.87321	0.62679
54	n-Octane	4.86841	125.70000	123.33554	2.36446
55	2,5-Dimethylhexane	4.77500	109.10000	112.35800	-3.25800
56	2-Methylheptane	4.82193	117.60000	117.87321	-0.27321
57	2,3,4-Trimethylpentane	4.72760	113.50000	106.78755	6.71245
58	2,3-Dimethylhexane	4.77500	115.60000	112.35800	3.24200
59	2,2,4-Trimethylpentane	4.72957	99.20000	107.01907	-7.81907
60	2,2-Dimethylhexane	4.77694	106.80000	112.58599	-5.78599
61	n-Heptane	4.62165	98.40000	94.33631	4.06369
62	2,2,3,3-tetramethylbutane	4.68370	106.50000	101.62842	4.87158
63	2,4-Dimethylpentane	4.52315	80.50000	82.76059	-2.26059
64	2-Methylhexane	4.57267	90.00000	88.58018	1.41982
65	2,2,3-Trimethylbutane	4.47516	80.90000	77.12080	3.77920
66	2,2-Dimethylpentane	4.52521	79.20000	83.00268	-3.80268
67	2,3-Dimethylbutane	4.25643	58.00000	51.41565	6.58435
68	2-methylpentane	4.30901	60.30000	57.59486	2.70514
69	2,4-Dimethylhexane	4.77500	109.40000	112.35800	-2.95800
70	3-Methylheptane	4.82193	118.90000	117.87321	1.02679

quality than ^{1}DCW modeling. Correlation weights, obtained in three probes of the Monte Carlo optimization, for calculation of the ^{2}DCW, are shown in Table 3. Calculation of the $^{2}DCW(a_k, {}^{1}EC_k)$ for m-xylene is shown in Table 4. The model of the hydrocarbon normal boiling points obtained in first probe of the optimization is the following

$$NBP = -448.8 + 117.52 * {}^{2}DCW(a_k, {}^{1}EC_k) \qquad (4)$$

$n = 70$, $r^2 = 0.9988$, $s = 5.86$, $F = 57619$ (training set)
$n = 70$, $r^2 = 0.9990$, $s = 4.82$, $F = 68071$ (test set)

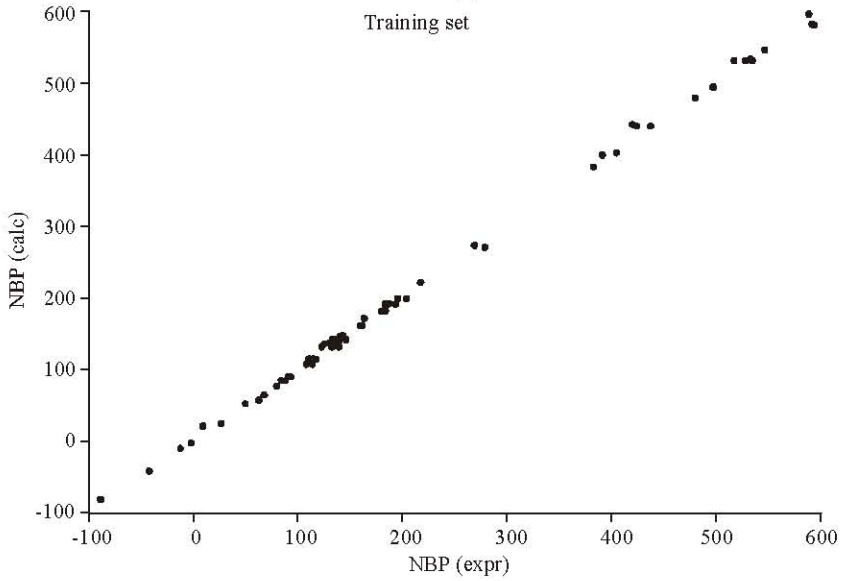


Fig. 1: Plot of experimental vs. calculated with Eq. (4) normal boiling points on the training set.

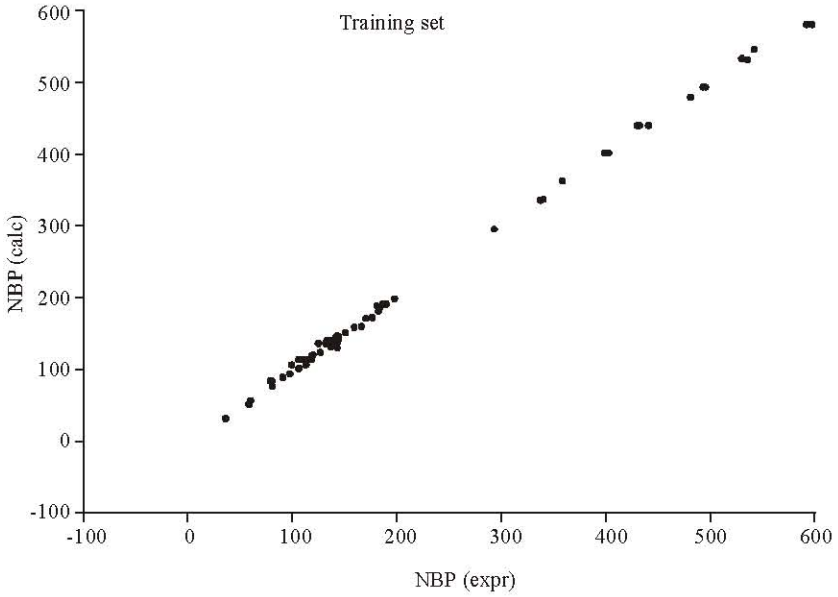


Fig. 2: Plot of experimental vs. calculated with Eq. (4) normal boiling points on the test set.

The results derived from this model is displayed graphically for training and test sets in Figures 1 and 2, respectively. This model is slightly better than one described in [11], but the it is considerably simpler than the previous one.

CONCLUSIONS

We have shown that utilization of the described cycle codes produce considerable improving statistical quality models of hydrocarbon normal boiling points regarding those calculation which do not take into account the presence of such cycles. In order to get a better conclusion on the goodness of the method proposed here, it is necessary to perform complementary calculations on other physicochemical properties and biological activities for other sets of molecules. Work along these lines are currently under development and results will be presented elsewhere in the forthcoming future.

REFERENCES

1. Chemoinformatics, J. Gasteiger and T. Engel, 2003. Editors, Wiley, New York, pp: 203-204.
2. QSPR/QSAR Studies by Molecular Descriptors, M. V. Diudea, Editor, Nova Science Publishers, Huntington, New York, 2001.
3. Castro, E.A., P.R. Duchowicz, A.A. Toropov, I.N. Nurgaliev and V. O. Kudyshkin, 2006. QSPR prediction of chain transfer constants employing optimal descriptors, Global J. Mol. Sci., 1: 33-41.
4. Duchowicz, P.R., E.A. Castro, A.A. Toropov and E. Benfenati, 2006. Applications of flexible molecular descriptors in the QSAR/QSPR study of heterocyclic drugs, in QSAR and Molecular Modeling Studies of Heterocyclic Drugs, S. Gupta, Editor, Vol.5, Springer-Verlag, Heidelberg, Germany.
5. Castro, E.A., A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanove, 2005. QSPR modeling of anti-HIV-1 activity of HEPT derivatives by optimization of correlation weights by Morgan extended connectivity in graph of atomic orbitals, J. Arg. Chem. Soc., 93: 109-121.
6. Marino, D.J.G., E.A. Castro and A.A. Toropov, 2006. Improved QSAR modeling of anti- HIV-1 activities by means of the optimized correlation weights of local graph invariants, Central Eur. J. Chem., 4: 135-148.
7. Castro, E.A., A.P. Toropova, A.A. Toropov and D.V. Mukhamedjanova, QSPR modeling of Gibbs free energy of organic compounds by weighting of nearest neighboring codes, Struct. Chem., 16: 311-330.
8. Pérez-González, M., A.A. Toropov, P.R. Duchowicz and E.A. Castro, 2004. QSPR calculation of normal boiling points of organic molecules based on the employment of correlation weighting of atomic orbitals with extended connectivity of zero- and first-order graphs of atomic orbitals, *Molecules,* 9: 1019-1033.
9. Castro, E.A., A.A. Toropov, A.I. Nesterova and A.U. Nazarov, 2004. QSPR modeling of enthalpies of formation from elements of coordination compounds by correlation weighting of nearest neughboring codes, J. Theor. Comp. Chem., 3: 31-41.
10. Duchowicz, P.R., E.A. Castro, A.A. Toropov, I.V. Nesterov and O.M. Nabiev, 2004. QSPR modeling the aqueous solubility of alcohols by optimization of correlation weights of local graph invariants, Mol. Divers., 8: 325-330.
11. Toropov, A.A., A.P. Toropova and I. Gutman, 2005. Comparison of QSPR models based on hydrogen-filled graphs and on graphs of atomic orbitals, Croat. Chem. Acta, 78: 503-509.
12. Toropov, A.A., A.P. Toropova, N.L. Voropaeva, I.N. Ruban and S.Sh. Rashidova, 1999. Using of matrix of maximal topological distances for QSPR modeling of normal boiling points of hydrocarbons which contents cycles, Russ. J. Struct. Chem., 40: 202-206.
13. Toropov, A.A., A.P. Toropova, D.V. Mukhamedzhanova and I. Gutman, 2005. Simplified molecular input line entry system (SMILES) as an alternative for constructing quantitative structure-property relationships (QSPR), Indian J. Chem., 44A, 1545-1552.
14. Toropov, A.A. and A.P. Toropova, 2003. QSPR modeling of alkane properties based on graph of atomic orbitals, J. Mol. Struct. (THEOCHEM), 637: 1-10.
15. Toropov, A.A. and E. Benfenati, 2006. QSAR models for Daphnia toxicity of pesticides based on combinations of topological parameters of molecular structures, Bioorg. Med. Chem., 14: 2779-2788.