CROATICA CHEMICA ACTA
CCACAA **78** (4) 503–509 (2005)

ISSN-0011-1643
CCA-3041
Original Scientific Paper

Comparison of QSPR Models Based on Hydrogen-Filled Graphs and on Graphs of Atomic Orbitals*

Andrey A. Toropov,[a] Alla P. Toropova,[a] and Ivan Gutman[b],**

[a]*Uzbekistan Academy of Science, Institute of Geology and Geophysics, A. Abdullaev street 17, 700125 Tashkent, Uzbekistan*

[b]*Faculty of Science, University of Kragujevac, P.O. Box 60, 34000 Kragujevac, Serbia & Montenegro*

RECEIVED JANUARY 18, 2005; REVISED APRIL 29, 2005; ACCEPTED MAY 3, 2005

Key words
chemical graph theory
molecular graph
graph of atomic orbitals
GAO
Morgan extended connectivity
QSPR models

QSPR models are studied for normal boiling points of alkanes, alkylbenzenes, and polyaromatic hydrocarbons, in terms of optimized correlation weights of local invariants of the hydrogen-filled graphs (HFGs) and of the graphs of atomic orbitals (GAOs). Morgan extended connectivities of the zeroth, first, and second order of the HFGs and GAOs were employed. The best QSPR model obtained is based on optimized correlation weights of the extended connectivity of the first order of the GAO. The statistical characteristics of this model are: $n = 70$, $r^2 = 0.9988$, $s = 5.8$ °C, $F = 57437$ (training set); $n = 70$, $r^2 = 0.9985$, $s = 6.7$ °C, $F = 45154$ (test set).

INTRODUCTION

Modeling of physical, chemical, pharmacological, toxicological and other properties of organic compounds *via* quantitative structure-property relationships (QSPR) represents an important field of activity in contemporary theoretical chemistry; for details see the book[1] and the chapters devoted to QSPR/QSPR studies in the edited books.[2–5] Recently, the so-called *optimization of correlation weights of local graph invariants* (OCWLI) has been developed and demonstrated to be a useful approach in QSPR studies.[6–11]

The OCWLI approach is based on a variation of the weights of the local graph invariants under consideration, aimed at obtaining an as high as possible correlation coefficient between the experimental and calculated values of the examined physico-chemical or pharmacological property. The respective weights varied by means of a Monte Carlo process, that is by adding to subtracting from them randomly generated increments. More details can be found in our earlier publications.[6,11]

In connection with the OCWLI method, it should be noted that it could be applied to any type of molecular graphs. Until now either the *hydrogen-filled molecular graphs* (HFGs) or the *graphs of atomic orbitals* (GAOs) were employed. Whereas the HFGs are well known in chemical graph theory,[12] the concept of GAOs was recently put forward by two of the present authors.[7,8,10]

The HGF is the molecular graph in which both the heavy and the hydrogen atoms are represented by vertices. Whereas Cayley, the inventor of the molecular-graph-concept, referred to these graphs as to »*plerograms*«,[13,14] in the more recent chemical literature the name »*hydrogen-filled molecular graph*« previals.[15–17] The molecular graph in which the hydrogen atoms are not represented

---

* Dedicated to Dr. Edward C. Kirby in happy celebration of his 70th birthday

** Author to whom correspondence should be addressed. (E-mail: gutman@knez.uis.kg.ac.yu) by vertices was originally[13,14] named »*kenogram*«, but nowadays is called either »*hydrogen-suppressed*« or »*hydrogen-depleted*« molecular graph.[12,15–17]

A rigorous graph-theoretical definition of the GAO can be found in the papers,[18,19] in which also some general mathematical properties thereof have been established.

In both the hydrogen-depleted and hydrogen-filled molecular graphs, vertices represent individual atoms. The idea behind the GAO is to represent by a vertex a group of atomic orbitals of the respective atom. These groups of atomic orbitals, for the most frequently encountered atoms in organic molecules, are the following:

atom	groups of atomic orbitals	n_i
H	$1s^1$	1
C	$1s^2\ 2s^2\ 2p^2$	3
N	$1s^2\ 2s^2\ 2p^3$	3
O	$1s^2\ 2s^2\ 2p^4$	3
F	$1s^2\ 2s^2\ 2p^5$	3
S	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^4$	5
Cl	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$	5
Br	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^5$	8
I	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 5s^2\ 5p^5$	11

Then the GAO, pertaining to a molecule M is defined as follows. Let G_H be the hydrogen-filled molecular graph of M. Let $\{v_1,v_2,...,v_N\}$ be the vertex set of G_H. Then the graph of atomic orbitals (GAO), pertaining to M, is obtained from G_H by replacing each of its vertex v_i, $i = 1,2,...,N$, by a set n_i distinct vertices, where the value of n_i depends on the type of the underlying atom, as specified in the above table. Vertices of GAO are adjacent if and only if they pertain to chemically bound atoms of M. Consequently, two vertices of a GAO, representing different groups of orbitals of the same atom, are not adjacent.

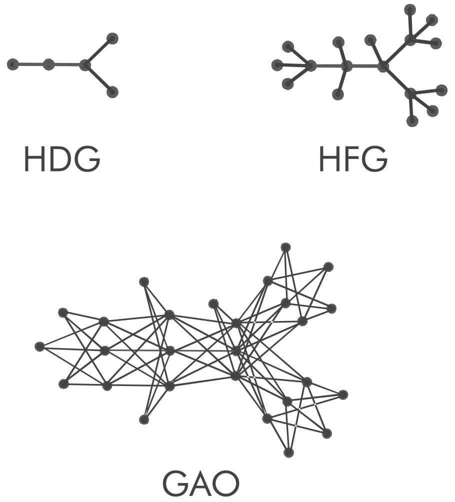

Figure 1. Molecular graphs of 2-methylbutane: HDG = hydrogen-depleted molecular graph, HFG = hydrogen-filled molecular graph, GAO = graph of atomic orbitals. Each vertex of HDG and HFG represents an atom; each vertex of GAO represents a group of atomic orbitals. For details see text and/or Ref. 18.

In Figure 1 are depicted the various molecular graphs of 2-methylbutane.

The GAO is constructed as tool for taking into account the electron configuration of various atoms contained in the molecule considered. Consequently, the GAO-concept was aimed at modeling the physical and chemical properties of organic compounds with heteroatoms.[6–11] In this work we examine the ability of the OCWLI-method for predicting a physico-chemical property, namely the normal boiling point, of compounds containing atoms of only two kinds, namely hydrogen and carbon. Boiling points of hydrocarbons are traditionally used for testing the correlation abilities of topological indices and their suitability as parameters in QSPR and QSAR studies.[12,20] However, in contrast to some earlier studies, we do not restrict our analysis to compounds belonging to the same structural class (*e. g.*, alkanes,[20,21] cycloalkanes,[22] planar benzenoids,[23] alkylbenzenes[24]), but consider a mixed set of all these types of hydrocarbons, see Table III.

TABLE I. Statistical characteristics of the OCWLI models based on different versions of the descriptors calculated by means of Eqs. (1) or (2). N^{op} is the number of optimized parameters; r, s, and F are the correlation coefficient, standard error, and Fischer F-ratio, respectively. The best model obtained is indicated by bold face.

Descriptor	N^{op}	Training set, n = 70			Test set, n = 70		
		r^2	s	F	r^2	s	F
$X(a,{}^{0}EC)$	5	0.9949	12.156	13183	0.9972	8.508	24127
$X(a,{}^{1}EC)$	10	0.9974	8.591	26460	0.9978	7.413	30609
$X(a,{}^{2}EC)$	23	0.9980	7.525	34506	0.9975	8.076	27136
$X(ao,{}^{0}EC)$	11	0.9953	11.635	14395	0.9973	8.203	25058
$\mathbf{X(ao,{}^{1}EC)}$	**30**	**0.9988**	**5.835**	**57437**	**0.9985**	**6.706**	**45154**
$X(ao,{}^{2}EC)$	87	0.9999	1.745	642746	0.9952	11.198	14204

TABLE II. The correlation weights needed for the calculation of $X(ao, {}^{1}EC)$. For details see text.

Invariant	Correlation weight
Atomic orbital, ao_k	
$1s^1$	–0.075
$1s^2$	0.235
$2s^2$	0.019
$2p^2$	0.016
First order extended connectivity, ${}^{1}EC_k$	
18	–0.014
21	2.579
24	1.425
27	–0.537
30	1.948
33	0.966
36	2.137
39	1.027
42	–0.723
45	1.060
48	0.736
51	1.222
54	1.150
57	1.313
60	1.088
63	1.278
66	1.371
69	1.591
72	1.722
75	1.596
78	2.039
81	1.869
84	2.753
87	2.335
90	3.227
96	3.249

METHOD

The generalized form of the descriptors used as tool for the present QSPR modeling is the following:

$$X(a, {}^{x}EC) = \sum_k C(a_k) + \sum_k C({}^{x}EC_k) \qquad (1)$$

$$X(ao, {}^{x}EC) = \sum_k C(ao_k) + \sum_k C({}^{x}EC_k) \qquad (2)$$

Formula (1) applies to HFGs whereas formula (2) to GAOs. In (1) *a* indicates that the summations go over **at**oms, whereas in (2) *ao* indicates that these summations go over groups of **a**tomic **o**rbitals. In Eq. (1), $C(a_k)$ is the correlation weight pertaining to a particular type of atom (H, C, O, *etc.*), represented by the *k*-th vertex of HFG. In Eq. (2), $C(ao_k)$ is the correlation weight of an atomic orbital ($1s^1$,…, $2p^3$, …, $3d^7$, …) in a given atom, represented by the *k*-th vertex of GAO. In both Eqs. (1) and (2) $C({}^{x}EC_k)$ stands for the correlation weight of the Morgan extended connectivity (*EC*) of the *x*-th order of the *k*-th vertex (of HFG or GAO). All summations in (1) and (2) embrace all the vertices of either the HGF or GAO.

RESULTS

Numerical values of the correlation weights were calculated by means of the above described and previously published[6,11] Monte Carlo optimization procedure. The data for the normal boiling points (see Table III) were taken from the paper.[25]

Table I contains results of the Monte Carlo optimization based on Morgan extended connectivity of orders 0, 1, and 2 of the HFG and GAO. From these data it is seen that the best result is obtained by means of GAOs, and by means of first-order extended connectivities ($x = 1$). Table II contains the correlation weights needed for the calculation of the $X(ao, {}^{1}EC)$-values, Eq. (2).

Our formula for computing the normal boiling points (NBPs) is:

$$\text{NBP} = -73.3 + 4.73\ X(ao, {}^{1}EC) \qquad (3)$$

Its quality is seen from the following statistical data (*cf.* Table I):

$$n = 70,\ r^2 = 0.9988,\ s = 5.8\ ^\circ\text{C},\ F = 57437$$
(training set)

$$n = 70,\ r^2 = 0.9985,\ s = 6.7\ ^\circ\text{C},\ F = 45154$$
(test set)

The experimental NBPs, as well as those calculated by means of Eq. (3) are shown in the Table III.

CONCLUSIONS

The vast majority of research done in chemical graph theory and, in particular, the vast majority of graph-invariant-based QSPR and QSAR studies employ hydrogen-depleted molecular graphs. Hydrogen-filled molecular graphs are used to a much lesser extent, whereas other types of molecular graphs are used only sporadically. On the other hand, there is no *a priori* reason why any of the existing types of molecular graphs should be preferred in attempts to design high-quality QSPR/QSAR models. It might well happen that non-standard molecular graphs would provide better QSPR/QSAR models that the traditional (hydrogen-depleted and hydrogen-filled) molecular graphs. In the present work we established such a case: within the class of QSPR models examined GAO performed better than the HFG.

TABLE III. Experimental normal boiling points (NBPs) of the compounds examined in this work, as well as the NBPs calculated by means of Eq. (3). For details see text.

ID	COMPOUNDS	$X(ao,{}^1EC)$	NBP / °C expr.	NBP / °C calc.	Expr.-calc.
	Training Set				
1	Ethane	–3.21	–88.60	–88.50	–0.10
2	Propane	6.60	–42.10	–42.06	–0.04
3	2,2-Dimethylpropane	18.17	9.50	12.63	–3.13
4	2-Methylbutane	21.47	27.80	28.25	–0.45
5	2-Methylpropane	15.34	–11.70	–0.75	–10.95
6	n-Butane	16.16	–0.50	3.14	–3.64
7	n-Hexane	29.00	68.70	63.87	4.83
8	3,3-Dimethylpentane	33.16	86.10	83.57	2.53
9	2,3-Dimethylpentane	32.43	89.80	80.08	9.72
10	3-Methylhexane	33.64	91.80	85.80	6.00
11	3-Ethylpentane	34.90	93.50	91.76	1.74
12	2,2,3-Trimethylpentane	38.23	109.80	107.54	2.26
13	2,3,3-Trimethylpentane	38.56	114.80	109.07	5.73
14	3,3-Dimethylhexane	39.19	112.00	112.08	–0.08
15	3-Ethyl-3-methylpentane	40.54	118.30	118.45	–0.15
16	3-Ethyl-2-methylpentane	39.48	115.60	113.44	2.16
17	2,2-Dimethylbutane	25.07	49.70	45.28	4.42
18	2,2,3,3-Tetramethylpentane	45.02	140.30	139.64	0.66
19	2,2,3,4-Tetramethylpentane	43.40	133.00	131.99	1.01
20	2,3,3-Trimethylhexane	44.59	137.70	137.59	0.11
21	2,2,4,4-Tetramethylpentane	40.95	122.30	120.42	1.88
22	2,2,4-Trimethylhexane	43.39	126.50	131.95	–5.45
23	2,4,4-Trimethylhexane	44.31	130.60	136.30	–5.70
24	3,3-Diethylpentane	46.56	146.20	146.92	–0.72
25	2,4-dimethyl–3-ethylpentane	44.65	136.70	137.88	–1.18
26	3-Ethyl-4-methylhexane	45.73	140.40	143.01	–2.61
27	4-Ethyl-2-methylhexane	44.81	133.80	138.63	–4.83
28	2-Methyloctane	45.72	143.30	142.95	0.35
29	3-Methyloctane	46.48	144.20	146.53	–2.33
30	4-Ethylheptane	44.88	141.20	138.99	2.21
31	2,2-Dimethylheptane	43.94	132.70	134.52	–1.82
32	2,5-Dimethylheptane	43.94	136.00	134.53	1.47
33	2,6-Dimethylheptane	43.18	135.20	130.95	4.25
34	3,5-Dimethylheptane	45.73	136.00	143.00	–7.00
35	3-Methyl-3-ethylhexane	46.57	140.60	146.97	–6.37
36	Benzene	35.72	80.10	95.66	–15.56
37	Toluene	40.69	110.60	119.15	–8.55
38	*m*-Xylene	45.44	139.10	141.63	–2.53
39	*p*-Xylene	45.65	138.40	142.63	–4.23
40	1-Methyl-3-ethylbenzene	49.41	161.30	160.41	0.89
41	1-Methyl-4-ethylbenzene	49.62	162.00	161.42	0.58
42	1,3,5-Trimethylbenzene	49.98	164.70	163.11	1.59
43	1,2,3,4-Tetramethylbenzene	58.18	205.00	201.87	3.13
44	1,3-Diethylbenzene	53.38	181.10	179.19	1.91
45	1,4-Diethylbenzene	53.59	183.80	180.20	3.60
46	1-Methyl-4-n-propylbenzene	55.07	183.80	187.20	–3.40
47	1,2-Dimethyl-3-ethylbenzene	56.45	193.90	193.72	0.18
48	1,3-Dimethyl-4-ethylbenzene	55.31	188.40	188.31	0.09

49	1,3-Dimethyl-5-ethylbenzene	53.95	183.80	181.89	1.91
50	1,2,4,5-Tetramethylbezene	56.62	196.80	194.50	2.30
51	Naphthalene	59.65	218.00	208.87	9.13
52	Acenaphthylene	71.54	270.00	265.10	4.90
53	Acenaphthene	73.27	279.00	273.25	5.75
54	Fluoranthene	93.60	383.00	369.43	13.57
55	Pyrene	95.34	393.00	377.66	15.34
56	Benzo(c)fluorene	100.28	406.00	401.03	4.97
57	Benzo(ghi)fluoranthene	106.29	422.00	429.46	–7.46
58	Benz(a)anthracene	106.16	425.00	428.84	–3.84
59	Dibenz(a,j)anthracene	129.16	531.00	537.61	–6.61
60	Cyclopenta(cd)pyrene	108.74	439.00	441.04	–2.04
61	Benzo(k)fluoranthene	117.11	481.00	480.63	0.37
62	Perylene	117.40	497.00	481.99	15.01
63	Anthanthrene	130.60	547.00	544.45	2.55
64	Indeno(1,2,3-cd)pyrene	129.07	534.00	537.22	–3.22
65	Dibenz(a,c)anthracene	128.22	535.00	533.16	1.84
66	Picene	128.64	519.00	535.17	–16.17
67	Coronene	142.78	590.00	602.04	–12.04
68	Dibenzo(a,i)pyrene	140.91	594.00	593.19	0.81
69	Dibenzo(a,l)pyrene	140.98	595.00	593.54	1.46
70	1,3-Dimethyl-2-ethylbenzene	56.35	190.00	193.24	–3.24
	TestSet				
1	3-Methylpentane	28.64	63.30	62.19	1.11
2	4-Methylheptane	38.63	117.70	109.42	8.28
3	n-Pentane	22.58	36.10	33.51	2.59
4	Dibenzo(a,e)pyrene	140.18	592.00	589.75	2.25
5	Dibenzo(a,h)pyrene	140.91	596.00	593.19	2.81
6	Dibenz(a,h)anthracene	129.16	535.00	537.61	–2.61
7	Benzo(ghi)perylene	130.09	542.00	542.01	–0.01
8	Indeno(1,2,3-cd)fluoranthene	127.55	531.00	529.99	1.01
9	Benz(a)pyrene	118.12	496.00	485.42	10.58
10	Benzo(e)pyrene	117.40	493.00	481.99	11.01
11	Naphthacene	106.68	440.00	431.27	8.73
12	Triphenylene	104.71	429.00	421.96	7.04
13	Benzo(b)fluoranthene	116.38	481.00	477.19	3.81
14	Chrysene	105.65	431.00	426.40	4.60
15	Benzo(a)fluorene	99.48	403.00	397.24	5.76
16	Benzo(b)fluorene	99.99	398.00	399.67	–1.67
17	Fluorene	76.48	294.00	288.47	5.53
18	4*H*-Cyclopenta(def)phenanthrene	89.17	359.00	348.50	10.50
19	Benzo(j)fluoranthene	117.40	480.00	481.99	–1.99
20	Phenanthrene	82.65	338.00	317.64	20.36
21	Anthracene	83.16	340.00	320.07	19.93
22	1,4-Dimethyl-2-ethylbenzene	55.31	186.90	188.31	–1.41
23	1,2,3,5-Tetramethylbenzene	56.82	198.20	195.46	2.74
24	1,2-Dimethyl-4-ethylbenzene	55.11	189.80	187.35	2.45
25	1-Methyl-2-n-propylbenzene	56.01	184.80	191.61	–6.81
26	1-Methyl-3-n-propylbenzene	54.86	181.80	186.20	–4.40
27	n-Butylbenzene	56.53	183.30	194.08	–10.78
28	1,2-Diethylbenzene	54.73	183.40	185.56	–2.16
29	1,2,3-Trimethylbenzene	52.28	176.10	173.98	2.12
30	1,2,4-Trimethylbenzene	51.13	169.40	168.57	0.83

31	n-Propylbenzene	50.11	159.20	163.71	–4.51
32	1-Methyl–2-ethylbenzene	50.55	165.20	165.82	–0.62
33	Ethylbenzene	44.66	136.20	137.93	–1.73
34	*o*-Xylene	46.38	144.40	146.08	–1.68
35	3,3,4-Trimethylhexane	45.61	140.50	142.43	–1.93
36	2,3,4-Trimethylhexane	43.26	139.00	131.33	7.67
37	3,3-Dimethylheptane	45.61	137.30	142.45	–5.15
38	3,4-Dimethylheptane	43.67	140.60	133.26	7.34
39	2,3-Dimethylheptane	43.84	140.50	134.06	6.44
40	2,4-Dimethylheptane	43.55	133.50	132.68	0.82
41	4-Methyloctane	45.05	142.50	139.78	2.72
42	3-Ethylheptane	46.31	143.00	145.73	–2.73
43	3,4-Dimethylhexane	38.68	117.70	109.65	8.05
44	n-Nonane	48.26	150.80	154.96	–4.16
45	2,3,5-Trimethylhexane	42.34	131.30	126.96	4.34
46	3-ethyl-2-methyl-hexane	44.47	138.00	137.05	0.95
47	2,2,5-Trimethylhexane	41.40	124.10	122.52	1.58
48	4,4-Dimethylheptane	45.22	135.20	140.60	–5.40
49	2,3-dimethyl-3-ethylpentane	44.57	142.00	137.54	4.46
50	2,2,3-Trimethylhexane	43.23	133.60	131.16	2.44
51	2,2-dimethyl-3-ethylpentane	45.87	133.80	143.66	–9.86
52	2,3,3,4-Tetramethylpentane	42.59	141.60	128.16	13.44
53	3-Ethylhexane	39.89	118.50	115.37	3.13
54	n-Octane	41.84	125.70	124.59	1.11
55	2,5-Dimethylhexane	36.76	109.10	100.59	8.51
56	2-Methylheptane	39.30	117.60	112.59	5.01
57	2,3,4-Trimethylpentane	37.01	113.50	101.76	11.74
58	2,3-Dimethylhexane	37.42	115.60	103.70	11.90
59	2,2,4-Trimethylpentane	36.22	99.20	98.01	1.19
60	2,2-Dimethylhexane	37.52	106.80	104.16	2.64
61	n-Heptane	35.42	98.40	94.23	4.17
62	2,2,3,3-tetramethylbutane	39.00	106.50	111.18	–4.68
63	2,4-Dimethylpentane	31.38	80.50	75.12	5.38
64	2-Methylhexane	32.88	90.00	82.23	7.77
65	2,2,3-Trimethylbutane	31.18	80.90	74.19	6.71
66	2,2-Dimethylpentane	31.10	79.20	73.79	5.41
67	2,3-Dimethylbutane	26.18	58.00	50.52	7.48
68	2-methylpentane	26.46	60.30	51.86	8.44
69	2,4-Dimethylhexane	38.55	109.40	109.06	0.34
70	3-Methylheptane	40.06	118.90	116.17	2.73

Therefore, the present work implies the following conclusions:

(i) In QSPR and QSAR studies one should not restrict the consideration to invariants of traditional molecular graphs, but – in justified cases – other types of molecular graphs should be taken into account.

(ii) The graph of atomic orbitals is found to be suitable for modeling such an ordinary physico-chemical property as the normal boiling point of hydrocarbons. If so, then GAO may prove to be useful also for subtler QSPR and QSAR applications.

(iii) The graph of atomic orbitals contains information not only on the structure of the respective organic molecule, but also on the electrons present in the atoms that form the respective molecule. When properly parametrized, its applicability may go far beyond the realm of hydrocarbons, thus overcoming one of the main limitations of traditional chemical graph theory.

(iv) The graph of atomic orbitals, viewed as a novel type of molecular graph, deserves to be investigated as such. Until now, very few general properties of GAOs have been established.[18,19]

REFERENCES

1. C. Hansch, A. Leo, and D. Hoeckman, *Exploiting QSAR*, Am. Chem. Soc., Washington, 1995.
2. W. Karcher and J. Devillers (Eds.), *Practical Applications of Quantitative Structure Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, EEC, Brussels, 1990.
3. J. Devillers (Ed.), *Comparative QSAR*, Taylor and Francis, Washington, 1998.
4. J. Devillers and A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, New York, 1999.
5. M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova, Huntington, 2001.
6. A. A. Toropov and A. P. Toropova, *Russ. J. Coord. Chem.* **24** (1998) 81–85.
7. A. A. Toropov and A. P. Toropova, *Russ. J. Coord. Chem.* **26** (2000) 398–405.
8. A. A. Toropov and A. P. Toropova, *Russ. J. Coord. Chem.* **27** (2001) 574–578.
9. A. A. Toropov and A. P. Toropova, *J. Mol. Struct. (Theochem)* **637** (2003) 1–10.
10. A. A. Toropov and A. P. Toropova, *J. Mol. Struct. (Theochem)* **711** (2004) 173–183.
11. A. A. Toropov and T. W. Schultz, *J. Chem. Inf. Comput. Sci.* **43** (2003) 560–567.
12. N. Trinajstić, *Chemical Graph Theory,* CRC Press, Boca Raton, 1982.
13. A. Cayley, *Philos. Mag.* **67** (1874) 444–446.
14. I. Gutman, D. Vidović, and L. Popović, *J. Chem. Soc., Faraday Trans.* **94** (1998) 852–860.
15. S. C. Basak and B. D. Gute, *SAR QSAR Environ. Res.* **7** (1997) 1–21.
16. S. C. Basak, A. T. Balaban, G. D. Grunwald, and B. D. Gute, *J. Chem. Inf. Comput. Sci.* **40** (2000) 891–898.
17. S. C. Basak, B. D. Gute, and A. T. Balaban, *Croat. Chem. Acta* **77** (2004) 331–341.
18. I. Gutman, A. A. Toropov, and A. P. Toropova, *MATCH–Commun. Math. Comput. Chem.* **53** (2005) 215–224.
19. I. Gutman, B. Furtula, A. A. Toropov, and A. P. Toropova, *MATCH–Commun. Math. Comput. Chem.* **53** (2005) 225–230.
20. Z. Mihalić and N. Trinajstić, *J. Chem. Educ.* **69** (1992) 701–712.
21. K. J. Burch, D. K. Wakefield, and E. G. Whitehead, *MATCH–Commun. Math. Comput. Chem.* **47** (2003) 25–52.
22. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **39** (1999) 788–802.
23. M. Firpo, L. Gavernet, E. A. Castro, and A. A. Toropov, *J. Mol. Struct. (Theochem)* **501/502** (2000) 419–425.
24. M. Randić, *New J. Chem.* **20** (1996) 1001–1009.
25. S. C. Basak, G. J. Niemi, and G. D. Veith, *J. Math. Chem.* **7** (1991) 243–272.

SAŽETAK

Usporedba QSPR modela zasnovanih na vodikom-popunjenim molekularnim grafovima i na grafovima atomskih orbitala

Andrey A. Toropov, Alla P. Toropova i Ivan Gutman

Istraživani su QSPR modeli za normalnu točku vrelišta alkana, alkilbenzena i poliaromatskih ugljikovodika, zasnovani na optimiziranim korelacijskim težinama lokalnih invarijanti vodikom-popunjenih molekularnih grafova (HFG) i grafova atomskih orbitala (GAO). Primjenjeni su Morganovi indeksi proširene povezanosti nultoga, prvoga i drugoga reda, kako za HFG tako i za GAO. Najbolji QSPR model je dobiven na osnovi optimiziranih korelacijskih težina za proširenu povezanost prvoga reda za GAO. Statističke karakteristike ovoga modela su: $n = 70$, $r^2 = 0.9988$, $s = 5.8$ °C, $F = 57437$ (training set); $n = 70$, $r^2 = 0.9985$, $s = 6.7$ °C, $F = 45154$ (test set).