

## Four New Topological Indices Based on the Molecular Path Code<sup>†</sup>

Alexandru T. Balaban,<sup>\*,‡</sup> Adrian Beteringhe,<sup>§</sup> Titus Constantinescu,<sup>§</sup> Petru A. Filip,<sup>||</sup> and Ovidiu Ivanciuc<sup>⊥</sup>

Texas A&M University at Galveston, 5007 Avenue U, Galveston, Texas 77551, Institute of Physical Chemistry “I. G. Murgulescu” of the Romanian Academy, Splaiul Independentei 202, Bucharest 77208, Romania, Center of Organic Chemistry “C. D. Nenitzescu” of the Romanian Academy, Splaiul Independentei 202A, Bucharest 77208, Romania, and Department of Biology and Molecular Biochemistry, University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555

Received November 12, 2006

The sequence of all paths  $p_i$  of lengths  $i = 1$  to the maximum possible length in a hydrogen-depleted molecular graph (which sequence is also called the molecular path code) contains significant information on the molecular topology, and as such it is a reasonable choice to be selected as the basis of topological indices (TIs). Four new (or five partly new) TIs with progressively improved performance (judged by correctly reflecting branching, centrality, and cyclicity of graphs, ordering of alkanes, and low degeneracy) have been explored. (i) By summing the squares of all numbers in the sequence one obtains  $\sum_i p_i^2$ , and by dividing this sum by one plus the cyclomatic number, a **Quadratic TI** is obtained:  $Q = \sum_i p_i^2 / (\mu + 1)$ . (ii) On summing the Square roots of all numbers in the sequence one obtains  $\sum_i p_i^{1/2}$ , and by dividing this sum by one plus the cyclomatic number, the TI denoted by  $S$  is obtained:  $S = \sum_i p_i^{1/2} / (\mu + 1)$ . (iii) On dividing terms in this sum by the corresponding topological distances, one obtains the **Distance-reduced index**  $D = \sum_i \{p_i^{1/2} / [i(\mu + 1)]\}$ . Two similar formulas define the next two indices, the first one with no square roots: (iv) distance-**Attenuated index**:  $A = \sum_i \{p_i / [i(\mu + 1)]\}$ ; and (v) the last TI with two square roots: **Path-count index**:  $P = \sum_i \{p_i^{1/2} / [i^{1/2}(\mu + 1)]\}$ . These five TIs are compared for their degeneracy, ordering of alkanes, and performance in QSPR (for all alkanes with 3–12 carbon atoms and for all possible chemical cyclic or acyclic graphs with 4–6 carbon atoms) in correlations with six physical properties and one chemical property.

### INTRODUCTION

After Wiener's<sup>1</sup> and Platt's indices (1947–1948),<sup>2</sup> Gordon and Scantlebury (1964)<sup>3</sup> and then Hosoya (1971) introduced their indices.<sup>4</sup> In 1975, Gutman, Rušćić, Trinajstić, and Wilcox published in 1975 the Zagreb indices.<sup>5</sup> In 1975, Randić described the molecular connectivity index  $\chi$ ,<sup>6</sup> and then during the same year, with Kier, Hall and Murray,<sup>7</sup> he developed the connectivity along longer paths. Rouvray's reviews (1971, 1973)<sup>8</sup> and the first Kier–Hall book (1976) appeared during the same period,<sup>9</sup> so that in 1979 only the above few topological indices had been described. After having edited in 1976 the first book on chemical applications of graph theory,<sup>10</sup> which did not include information on topological indices, one of us (A.T.B.) published in 1979 a paper introducing five new “centric topological indices” for acyclic graphs.<sup>11</sup> These indices were cited in both editions of Trinajstić's book.<sup>12</sup> Hundreds of TIs have been described so far.<sup>13–15</sup> They have the advantage that, unlike other molecular descriptors, they can be computed rapidly for any known or unknown chemical structure.

The present article describes another set of five topological indices, all derived from the sequence of path counts, with the first of them ( $Q$ ) inspired by the Zagreb authors' way of converting several sequences of integers with identical sums for isomeric molecules into distinct molecular descriptors.

Randić and co-workers described the *molecular path code* as the ordered sequence of path counts starting with  $p_1$ , the number of paths of length 1 (i.e. edges).<sup>16</sup> They also developed an algorithm (ALLPATH),<sup>17</sup> mentioned by Jurs,<sup>18</sup> for computing the number of paths of all possible lengths, noting that for polycyclic graphs these numbers are hard to find “by hand”. The program DRAGON made available by Todeschini also allows the computation of path sequences.<sup>19</sup> Observing that all path sequences for constitutional isomers of alkanes with  $n = 3–8$  carbon atoms are different, Randić called such sequences “God-given names” of alkanes.<sup>20</sup> However, starting with nonane isomers, the molecular path codes do present degeneracies, as indicated in Table 1 and as will be discussed in the next section. If the order of paths and clusters in the sequence differs, even if the numbers in the sequence are the same, we consider that the sequences are different (not degenerate). A sequence of numbers is cumbersome to work with in characterizing molecular structures.

In searching for ways to concentrate the structural information contained in the molecular path code into a more compact molecular descriptor, Randić showed that on using only the numbers of paths of lengths 2 and 3 one may find

<sup>†</sup> Dedicated to Professor Nenad Trinajstić on the occasion of his 70th birthday.

\* Corresponding author e-mail: balabana@tamug.edu.

<sup>‡</sup> Texas A&M University at Galveston.

<sup>§</sup> Institute of Physical Chemistry “I. G. Murgulescu” of the Romanian Academy.

<sup>||</sup> Center of Organic Chemistry “C. D. Nenitzescu” of the Romanian Academy.

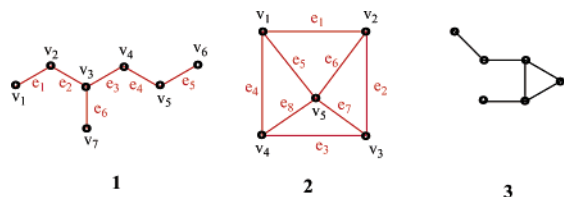
<sup>⊥</sup> University of Texas Medical Branch.

**Table 1.** Alkanes C<sub>3</sub>–C<sub>10</sub> with their Molecular Path Codes, TIs, and Physical Properties<sup>a</sup>

no.	alkanes (sorted by S)	abbreviation	molecular path code	n	W	J	Q	S	D	R	P	C <sub>p</sub> (J/ K·mol)	ρ (kg/ m <sup>3</sup> )	Δ <sub>r</sub> G <sup>0</sup> <sub>300</sub> (g) (kJ/mol)	ΔH <sub>vap</sub> <sup>300</sup> (kJ/mol)	NBP (°C) (obs)	NBP (°C) (calc)	resid	
1	propane	C3	2,1	3	4	1.6334	5	2.4142	1.9142	2.5000	2.12132								
2	butane	C4	3,2,1	4	10	1.9747	14	4.1462	2.7724	4.3333	3.30940								
3	2-methylpropane	2-M-C3	3,3	4	9	2.3237	18	3.4641	2.5980	4.5000	2.95679								
4	pentane	C5	4,3,2,1	5	20	2.1906	30	6.1462	3.5874	6.4167	4.45124								
5	2-methylbutane	2-M-C4	4,4,2	5	18	2.5395	36	5.4142	3.4714	6.6667	4.23071								
6	2,2-dimethylpropane	2,2-MM-C3	4,6	5	16	3.0237	52	4.4494	3.2247	7.0000	3.73205								
7	hexane	C6	5,4,3,2,1	6	35	2.3390	55	8.3823	4.3669	8.7000	5.80460								
8	2-methylpentane	2-M-C5	5,5,3,2	6	32	2.6272	63	7.6184	4.2850	9.0000	5.52431	143.01	648.52	1.3687	-4.05	26.61	60.9	63.7	-2.8
9	3-methylpentane	3-M-C5	5,5,4,1	6	31	2.7541	67	7.4721	4.2707	9.0833	5.47190	140.88	659.76	1.3739	-2.12	26.32	63.3	65.4	-2.1
10	2,3-dimethylbutane	2,3-MM-C4	5,6,4	6	29	3.1684	77	6.6855	4.1274	9.3333	5.12281	140.21	657.02	1.3723	-1.77	24.77	58.9	68.3	-9.4
11	2,2-dimethylpentane	2,2-MM-C4	5,7,3	6	28	2.9935	83	6.6138	4.1362	9.5000	5.10689	142.26	644.46	1.3660	-7.42	25.40	58.1	63.8	-5.7
12	heptane	C7	6,5,4,3,2,1	7	56	2.4475	91	10.8318	5.1167	11.1500	7.09205	166.00	679.50	1.3851	9.50	33.56	98.4	87.7	10.7
13	2-methylhexane	2-M-C6	6,6,4,3,2	7	52	2.6783	101	10.0452	5.0567	11.4833	6.83472	165.40	674.34	1.3823	4.90	31.21	90.1	86.9	3.2
14	3-methylhexane	3-M-C6	6,6,5,3,1	7	50	2.8318	107	9.8670	5.0526	11.6167	6.78577	164.50	682.88	1.3861	6.60	30.71	91.9	89.2	2.7
15	2,4-dimethylpentane	2,4-MM-C5	6,7,4,4	7	48	2.9532	117	9.0952	4.9390	12.7500	6.48502	171.70	668.23	1.3788	4.90	29.58	79.2	85.7	-6.5
16	3-ethylpentane	3-E-C5	6,6,6,3	7	48	2.9923	117	9.0805	4.9237	11.8333	6.47501	166.80	693.92	1.3911	12.70	31.71	89.8	86.5	3.3
17	2,3-dimethylpentane	2,3-MM-C5	6,7,6,2	7	46	3.1442	125	8.9589	4.9424	12.0833	6.47021	161.80	690.81	1.3895	7.60	28.62	80.5	89.3	-8.8
18	2,2-dimethylhexane	2,2-MM-C5	6,8,4,3	7	46	3.1545	125	9.0099	4.9633	11.7500	6.46178	167.70	669.48	1.3800	2.10	29.50	86.0	89.7	-3.7
19	3,3-dimethylpentane	3,3-MM-C5	6,8,6,1	7	44	3.3604	137	8.7274	4.9302	12.0000	6.44163	166.70	689.16	1.3884	4.80	29.33	93.5	92.8	0.7
20	2,2,3-trimethylbutane	2,2,3-MMM-C4	6,9,6	7	42	3.5412	153	7.8989	4.7659	12.5000	5.98502	164.20	685.64	1.3869	6.30	28.28	80.9	89.3	-8.4
21	octane	C8	7,6,5,4,3,2,1	8	84	2.5301	140	13.4775	5.8408	13.7429	8.39870	188.70	698.54	1.3951	17.67	38.12	125.7	112.9	12.8
22	2-methylheptane	2-M-C7	7,7,5,4,3,2	8	79	2.7158	152	12.6738	5.7960	14.1000	8.15952	188.20	693.87	1.3926	13.37	35.82	117.7	111.2	6.5
23	3-methylheptane	3-M-C7	7,7,6,4,3,1	8	76	2.8621	160	12.4730	5.7982	14.2667	8.11363	186.82	701.73	1.3961	13.79	35.31	117.0	113.1	3.9
24	4-methylheptane	4-M-C7	7,7,6,5,2,1	8	75	2.9196	164	12.3912	5.7936	14.3167	8.08953	188.03	700.71	1.3955	17.40	35.06	117.7	113.8	3.9
25	2,5-dimethylhexane	2,5-MM-C6	7,8,5,4,4	8	74	2.9278	170	11.7102	5.7053	14.7667	7.83266	186.52	689.37	1.3900	11.40	33.39	106.8	108.6	-1.8
26	3-ethylhexane	3-E-C6	7,7,7,5,2	8	72	3.0744	176	11.5875	5.6924	14.4667	7.83117	190.58	709.45	1.3992	15.53	36.07	114.0	111.2	2.8
27	2,4-dimethylhexane	2,4-MM-C6	7,8,6,5,2	8	71	3.0988	178	11.5739	5.7183	14.6500	7.81045	193.35	696.17	1.3929	13.07	33.76	109.0	111.6	-2.6
28	2,2-dimethylhexane	2,2-MM-C6	7,9,5,4,3	8	71	3.1118	180	11.6138	5.7375	14.7333	7.80573	189.33	691.11	1.3910	12.15	34.23	109.0	112.1	-3.1
29	2,3-dimethylhexane	2,3-MM-C6	7,8,7,4,2	8	70	3.1708	182	11.5341	5.7247	14.4833	7.79459	185.18	708.16	1.3988	17.20	33.05	112.0	112.9	-0.9
30	3,4-dimethylhexane	3,4-MM-C6	7,8,8,4,1	8	68	3.2925	194	11.3026	5.7027	15.0333	7.74181	182.72	715.15	1.4018	18.43	32.47	117.7	114.1	3.6
31	3,3-dimethylhexane	3-E-MM-C5	7,9,7,4,1	8	67	3.3549	196	11.2915	5.7276	14.8667	7.72595	191.96	705.95	1.3978	15.13	33.43	109.8	115.3	-5.5
32	3-ethyl-2-methylpentane	3,3-MM-C6	7,8,8,5	8	67	3.3734	202	10.5386	5.5617	15.1667	7.40006	193.05	715.20	1.4017	20.68	34.31	118.0	109.4	8.6
33	2,2,4-trimethylpentane	2,2,4-C5	7,10,5,6	8	66	3.3889	210	10.4935	5.5846	15.1667	7.39752	189.45	687.84	1.3890	15.70	32.55	99.2	109.5	-10.3
34	2,3,4-trimethylpentane	2,3,4-C5	7,9,8,4	8	65	3.4642	210	10.4741	5.5885	14.9167	7.39677	192.72	715.09	1.4020	20.76	32.55	114.8	111.1	3.7
35	3-ethyl-3-methylpentane	3-E3M-C5	7,9,9,3	8	64	3.5832	220	10.3778	5.5787	15.4167	7.38083	189.07	723.54	1.4055	24.36	33.26	113.5	113.2	0.3
36	2,2,3-trimethylpentane	2,2,3-MMM-C5	7,10,8,3	8	63	3.6233	222	10.3685	5.6027	15.2500	7.36514	186.77	712.03	1.4007	19.45	32.13	115.6	113.9	1.7
37	2,3,3-trimethylpentane	2,3,3-MMM-C5	7,10,9,2	8	62	3.7083	234	10.2222	5.5804	15.5000	7.32097	188.20	722.30	1.4052	20.04	32.17	118.3	114.8	3.5
38	2,2,3,3-tetramethylbutane	2,2,3,3-MMMM-C4	7,12,9	8	58	4.0204	274	9.1098	5.3778	16.0000	6.82729	188.28			24.04	31.84	106.5	112.6	-6.1
39	nonane	C9	8,7,6,5,4,3,2,1	9	120	2.5951	204	16.3060	6.5424	16.4607	9.72111								
40	2-methyloctane	2-M-C8	8,8,6,5,4,3,2	9	114	2.7467	218	15.4886	6.5088	16.8357	9.49673	210.90	709.60	1.4008	21.60	40.42	143.0	136.3	6.7
41	3-methyloctane	3-M-C8	8,8,7,5,4,3,1	9	110	2.8766	228	15.2707	6.5151	17.0262	9.45348	209.70	716.70	1.4040	22.00	39.92	143.9	137.8	6.1
42	4-methyloctane	4-M-C8	8,8,7,6,4,2,1	9	108	2.9147	234	15.1663	6.5154	17.1095	9.43043	210.40	716.30	1.4041	21.00	39.71	131.9	137.9	-6.0
43	2,6-dimethylheptane	2,6-MM-C7	8,9,6,5,4,4	9	108	2.9548	238	14.5139	6.4372	17.5500	9.19827	210.40	704.50	1.3985	19.80	37.99	141.9	133.7	8.2
44	2,5-dimethylheptane	2,5-MM-C7	8,9,7,5,5,2	9	104	3.0608	248	14.3605	6.4522	17.2167	9.19291	208.20	713.60	1.4015	18.20	37.53	139.9	135.3	4.6
45	3-ethylheptane	3-E-C7	8,8,8,6,4,2	9	104	3.0730	248	14.3489	6.4335	17.4167	9.17265	213.00	722.50	1.4070	26.40	40.71	134.9	135.6	-0.7
46	2,2-dimethylheptane	2,2-MM-C7	8,10,6,5,4,3	9	104	3.0923	250	14.4083	6.4737	17.5500	9.17255	212.40	706.60	1.3995	19.50	38.83	132.9	136.1	-3.2
47	2,3-dimethylheptane	2,3-MM-C7	8,9,8,5,4,2	9	102	3.1513	254	14.3071	6.4659	17.3000	9.15794	207.70	722.00	1.4064	23.50	37.82	136.9	136.8	0.1
48	2,4-dimethylheptane	2,4-MM-C7	8,9,7,7,3,2	9	102	3.1553	256	14.2661	6.4538	17.5167	9.15209	217.10	711.50	1.4011	20.80	38.16	134.9	136.7	-1.8
49	4-ethylheptane	4-E-C7	8,8,8,7,4,1	9	102	3.1753	258	14.1310	6.4135	17.8833	9.11819	214.30	722.30	1.4067	26.80	40.50	140.5	136.2	4.3
50	3,5-dimethylheptane	3,5-MM-C7	8,9,8,6,4,1	9	100	3.2321	262	14.1063	6.4502	17.6333	9.11016	214.60	716.60	1.4046	22.00	38.07	135.6	137.1	-1.5
51	2,2,5-trimethylhexane	2,2,5-MMM-C6	8,11,6,5,6	9	98	3.2807	282	13.2801	6.3521	17.7667	9.08938	209.10	703.20	1.3973	15.30	36.36	134.9	133.7	1.2
52	3,3-dimethylheptane	3,3-MM-C7	8,10,8,5,4,1	9	98	3.3074	270	14.0552	6.4780	17.3833	9.08697	214.00	721.60	1.4063	22.00	38.20	133.1	138.5	-5.4
53	3,4-dimethylheptane	3,4-MM-C7	8,9,9,6,3,1	9	98	3.3248	272	14.0099	6.4538	17.9833	9.06106	206.00	727.50	1.4091	24.90	37.02	141.3	138.5	2.8
54	4-ethyl-2-methylhexane	4-E-2-M-C6	8,9,8,7,4	9	98	3.3301	274	13.3026	6.3326	17.9667	8.81661	219.70	724.20	1.4054	24.50	39.25	143.3	133.1	10.2
55	2,3,5-trimethylhexane	2,3,5-MMM-C6	8,10,8,6,4	9	96	3.3766	280	13.2686	6.3647	17.9500	8.80135	212.50	717.90	1.4037	22.20	36.02	126.5	133.9	-7.4
56	4,4-dimethylheptane	4,4-MM-C7	8,10,8,7,2,1	9	96	3.4101	282	13.8791	6.4633	17.7167	8.80001	217.20	718.30	1.4053	25.80	37.53	137.1	137.9	-0.8
57	3-ethyl-2-methylhexane	3-E2M-C6	8,9,9,7,3	9	96	3.3911	284	13.2062	6.3362	18.1833	8.79863	216.10	729.00	1.4091	26.40	38.70	123.9	134.8	-10.9
58	2,2,4-trimethylhexane	2,2,4-MMM-C6	8,11,7,7,3	9	94	3.4673	292	13.1686	6.3765	18.3500	8.79831	210.70	711.80	1.4014	23.60	36.61	138.9	135.4	3.5
59	3-ethyl-4-methylhexane	3-E4M-C6	8,9,10,7,2	9	94	3.4995	298	13.0506	6.3268	17.8500	8.77927	215.20	738.00	1.4128	29.90	38.07	130.9	135.4	-4.5
60	2,2,3-trimethylhexane	2,2,3-MMM-C6	8,11,9,5,3	9	92	3.5758	300	13.1131	6.3921	18.3167	8.76195	209.90	725.70	1.4082	27.20				

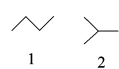
Table 1. (Continued)

no.	alkanes (sorted by S)	abbreviation	molecular path code	<i>n</i>	<i>W</i>	<i>J</i>	<i>Q</i>	<i>S</i>	<i>D</i>	<i>R</i>	<i>P</i>	<i>C<sub>p</sub></i> (J/(K·mol))	<i>ρ</i> (kg/ m <sup>3</sup> )	<i>n<sub>D</sub><sup>25</sup></i>	$\Delta_f G_{300}^0$ (g) (kJ/mol)	$\Delta H_{vap}^{300}$ (kJ/mol)	NBP (°C) (obs)	NBP (°C) (calc)	resid
79	5-methylnonane	5-M-C9	9,9,8,7,6,3,2,1	10	149	2.9897	325	18.0699	7.2098	20.4286	10.56863						165.1	163.2	1.9
80	2,6-dimethyloctane	2,6-MM-C8	9,10,8,6,5,5,2	10	146	3.0335	335	17.3265	7.1582	20.0714	10.56076	231.90	723.60	1.4084	26.90	42.09	158.9	160.1	-1.2
81	2,2-dimethyloctane	2,2-MM-C8	9,11,7,6,5,4,3	10	146	3.0478	<b>337</b>	17.3798	7.1805	20.4524	10.54388	235.10	720.80	1.4060	27.70	43.43	154.0	160.6	-6.6
82	3-ethylheptane	3-E-C8	9,9,9,7,5,4,2	10	145	3.0941	<b>337</b>	17.2960	7.1440	20.2857	10.54120	235.80	735.90	1.4136	34.90	45.31	166.0	161.2	4.8
83	2,3-dimethyloctane	2,3-MM-C8	9,10,9,6,5,4,2	10	143	3.1378	<b>343</b>	17.2620	7.1760	20.4024	10.52901	230.50	734.40	1.4217	32.00	42.43	164.2	161.9	2.3
84	2,5-dimethyloctane	2,5-MM-C8	9,10,8,7,6,3,2	10	143	3.1278	<b>343</b>	17.2322	7.1659	20.2024	10.52727	231.80	726.40	1.4089	26.90	41.92	156.9	161.5	-4.6
85	2,4-dimethyloctane	2,4-MM-C8	9,10,8,8,5,3,2	10	142	3.1644	347	17.2014	7.1689	20.4524	10.52490	239.40	722.60	1.4069	28.80	42.76	155.9	162.2	-6.3
86	3,6-dimethyloctane	3,6-MM-C8	9,10,9,6,6,4,1	10	141	3.1778	351	17.0612	7.1596	20.8095	10.49646	229.60	732.90	1.4115	28.90	41.63	160.0	161.5	-1.5
87	4-ethyloctane	4-E-C8	9,9,9,8,6,3,1	10	141	3.2157	353	17.0099	7.1285	20.5095	10.48277	236.50	734.30	1.4131	33.40	45.10	163.4	162.1	1.3
88	3,3-dimethyloctane	3,3-MM-C8	9,11,9,6,5,4,1	10	138	3.2887	<b>361</b>	17.0021	7.1940	20.7262	10.46976	237.10	735.10	1.4142	30.50	42.80	161.2	163.6	-2.4
89	3,5-dimethyloctane	3,5-MM-C8	9,10,9,8,5,3,1	10	138	3.2787	<b>361</b>	16.9588	7.1669	20.6429	10.46740	238.30	732.90	1.4115	29.10	42.47	158.9	163.1	-4.2
90	5-ethyl-2-methylheptane	5-E-2-M-C7	9,10,9,7,6,4	10	138	3.2658	363	16.2575	7.0658	20.3429	10.44810	234.30	731.80	1.4111	31.20	42.93	159.7	158.9	0.8
91	3,4-dimethyloctane	3,4-MM-C8	9,10,10,7,5,3,1	10	137	3.3104	365	16.9384	7.1754	20.9762	10.44679	229.30	741.80	1.4159	33.00	41.80	166.0	163.6	2.4
92	4-propylheptane	4-P-C7	9,9,9,9,6,3	10	138	3.3078	369	16.1815	7.0285	20.9765	10.44673	237.70	732.10	1.4113	38.20	44.85	161.9	159.4	2.5
93	2,2,6-trimethylheptane	2,2,6-MMM-C7	9,12,7,6,5,6	10	139	3.2176	<b>371</b>	16.2449	7.0818	20.8095	10.43133	234.80	723.80	1.4178	24.20	40.96	148.3	157.6	-9.3
94	2,3,6-trimethylheptane	2,3,6-MMM-C7	9,11,9,6,6,4	10	136	3.3001	<b>371</b>	16.2145	7.0939	20.8667	10.21395	228.50	734.70	1.4131	28.50	39.75	149.9	157.8	-7.9
95	4,5-dimethyloctane	4,5-MM-C8	9,10,10,8,5,2,1	10	135	3.3845	375	16.8032	7.1681	21.3333	10.20708	230.10	743.20	1.4167	35.30	41.51	162.4	162.9	-0.5
96	4,4-dimethyloctane	4,4-MM-C8	9,11,9,8,5,2,1	10	134	3.4207	377	16.7953	7.1911	20.6167	10.20295	239.30	731.20	1.4122	31.90	42.30	157.5	163.6	-6.1
97	2,4,6-trimethylheptane	2,4,6-MMM-C7	9,11,8,9,4,4	10	135	3.3444	<b>379</b>	16.1450	7.0844	20.8333	10.20176	246.30	719.00	1.4071	28.40	41.13	146.8	158.4	-11.6
98	3-ethyl-2-methylheptane	3-E-2-M-C7	9,10,10,8,5,3	10	134	3.4010	<b>379</b>	16.1211	7.0782	21.0667	10.19755	238.50	739.80	1.4151	35.70	43.30	166.0	159.5	6.5
99	2,2,5-trimethylheptane	2,2,5-MMM-C7	9,12,8,6,7,3	10	134	3.3621	<b>383</b>	16.1198	7.1050	21.2167	10.18402	230.50	728.10	1.4104	23.80	40.50	146.0	158.5	-12.5
100	4-ethyl-2-methylheptane	4-E-2-M-C7	9,10,9,9,6,2	10	134	3.3897	<b>383</b>	16.0259	7.0567	20.8333	10.18313	243.40	732.20	1.4114	31.60	43.64	160.0	158.6	1.4
101	5-ethyl-3-methylheptane	5-E-3-M-C7	9,10,10,8,6,2	10	133	3.4187	385	16.0166	7.0679	21.2333	10.16685	240.90	736.80	1.4141	33.10	43.35	158.3	159.1	-0.8
102	2,3,5-trimethylheptane	2,3,5-MMM-C7	9,11,10,7,6,2	10	131	3.4645	391	15.9883	7.0994	21.1167	10.16662	233.90	745.10	1.4169	30.30	40.12	157.0	160.1	-3.1
103	2,5,5-trimethylheptane	2,5,5-MMM-C7	9,12,9,6,7,2	10	131	3.4755	<b>395</b>	15.9735	7.1092	21.5000	10.16644	234.20	736.20	1.4149	25.80	40.54	152.8	160.2	-7.4
104	2,2,3-trimethylheptane	2,2,3-MMM-C7	9,12,10,6,5,3	10	130	3.5278	<b>395</b>	16.0439	7.1344	21.1667	10.16251	232.50	738.50	1.4145	34.80	41.30	163.9	161.7	-3.7
105	2,4,5-trimethylheptane	2,4,5-MMM-C7	9,11,10,8,5,2	10	130	3.5045	<b>395</b>	15.9576	7.1024	21.2500	10.16029	234.10	737.30	1.4160	36.10	39.98	157.0	160.7	-3.7
106	2,2,4-trimethylheptane	2,2,4-MMM-C7	9,12,8,9,4,3	10	131	3.4704	<b>395</b>	16.0245	7.1135	20.4500	10.15591	234.70	725.70	1.4092	31.90	41.05	148.9	160.3	-11.4
107	4-isopropylheptane	4-I-C7	9,10,10,9,6,1	10	131	3.5000	399	15.7740	7.0417	20.8667	10.14882	239.20	735.40	1.4132	37.90	43.10	158.8	159.5	-0.7
108	3-ethyl-3-methylheptane	3-E-3-M-C7	9,11,11,7,5,2	10	129	3.5878	401	15.9292	7.1082	21.3000	10.14605	236.20	746.30	1.4185	38.20	42.13	163.9	162.4	1.5
109	3-ethyl-4-methylheptane	3-E-4-M-C7	9,10,11,9,4,2	10	130	3.5377	403	15.8931	7.0723	20.7833	10.14091	236.20	746.60	1.4183	36.90	42.47	167.0	160.9	6.1
110	2,3,4-trimethylheptane	2,3,4-MMM-C7	9,11,11,8,4,2	10	128	3.5845	407	15.8758	7.1066	21.0500	10.12270	237.60	748.50	1.4195	37.20	40.96	155.9	161.9	-6.0
111	4-ethyl-3-methylheptane	4-E-3-M-C7	9,10,11,9,5,1	10	129	3.5668	409	15.7149	7.0505	21.4333	10.11463	238.60	746.80	1.4184	38.60	42.51	167.0	160.5	6.5
112	2,3,3-trimethylheptane	2,3,3-MMM-C7	9,12,11,6,5,2	10	127	3.6366	411	15.8805	7.1328	21.5000	10.09769	235.10	748.80	1.4202	37.30	41.00	160.1	162.9	-2.8
113	3,3,5-trimethylheptane	3,3,5-MMM-C7	9,12,10,8,5,1	10	126	3.6478	415	15.6908	7.1071	21.7167	10.07504	234.10	739.00	1.4170	34.10	40.46	155.7	161.9	-6.2
114	2,4,4-trimethylheptane	2,4,4-MMM-C7	9,12,9,10,3,2	10	127	3.6333	<b>419</b>	15.7726	7.1047	20.9500	10.06550	238.90	734.60	1.4143	35.90	40.54	149.6	162.0	-12.4
115	2,5-dimethyl-3-ethylhexane	2,5-MM-3-E-C6	9,11,10,9,6	10	127	3.6001	<b>419</b>	14.9283	6.9523	21.3833	10.06274	240.80	736.80	1.4232	33.60	41.34	157.0	156.6	0.4
116	4-ethyl-4-methylheptane	4-E-4-M-C7	9,11,11,9,4,1	10	126	3.6905	421	15.6332	7.0805	21.4667	10.06210	239.20	747.20	1.4187	40.30	41.46	159.0	162.5	-3.5
117	2,2-dimethyl-4-ethylhexane	2,2-MM-4-E-C6	9,12,9,9,6	10	126	3.6345	423	14.9135	6.9719	21.0833	10.05917	236.10	730.20	1.4107	36.10	41.92	147.1	157.3	-10.2
118	3,4,5-trimethylheptane	3,4,5-MMM-C7	9,11,12,8,4,1	10	125	3.5998	427	15.6091	7.0867	21.7667	10.04655	235.10	751.90	1.4229	39.70	41.14	164.0	160.2	3.8
119	2,2,4,5-tetramethylhexane	2,2,4,5-MMMM-C6	9,13,9,8,6	10	124	3.6878	431	14.8834	6.9997	21.7833	9.79357	229.20	731.61	1.4132	32.80	38.83	147.9	156.6	-8.7
120	3,3,4-trimethylheptane	3,3,4-MMM-C7	9,12,12,7,4,1	10	123	3.7821	<b>435</b>	15.5739	7.1148	21.7000	9.79121	233.60	752.70	1.4236	38.70	40.46	164.0	162.5	7.4
121	2,2,3,5-tetramethylhexane	2,2,3,5-MMMM-C6	9,13,10,7,6	10	123	3.7445	<b>435</b>	14.8630	7.0082	21.4500	9.77698	235.80	733.60	1.4119	32.10	39.54	148.7	157.6	-8.9
122	3-isopropyl-2-methylhexane	3-I-2-M-C6	9,11,11,10,4	10	124	3.7298	<b>439</b>	14.7955	6.9544	21.9667	9.77306	231.80	743.60	1.4172	46.80	40.46	163.0	156.9	6.1
123	3,4-diethylhexane	3,4-EE-C6	9,10,12,10,4	10	125	3.7009	441	14.7886	6.9264	21.2833	9.76639	246.90	747.20	1.4167	45.00	43.68	162.8	156.2	6.6
124	3,4,4-trimethylheptane	3,4,4-MMM-C7	9,12,12,8,3,1	10	122	3.8277	<b>443</b>	15.4886	7.1069	21.7165	9.75873	235.60	753.50	1.4235	40.30	40.08	164.0	162.8	1.2
125	2,3-dimethyl-4-ethylhexane	2,3-MM-4-E-C6	9,11,12,9,4	10	123	3.7602	<b>443</b>	<b>14.7807</b>	6.9630	21.7167	9.75877	243.00	751.60	1.4203	42.10	42.43	164.0	157.4	6.6
126	2,2-dimethyl-3-ethylhexane	2,2-MM-3-E-C6	<b>9,12,11,9,4</b>	10	122	3.8091	<b>443</b>	<b>14.7807</b>	6.9875	21.8000	9.75815	227.70	744.70	1.4174	44.40	40.50	159.0	158.6	0.4
127	2,4-dimethyl-4-ethylhexane	2,4-MM-4-E-C6	<b>9,12,11,9,4</b>	10	122	3.8035	<b>443</b>	<b>14.7807</b>	6.9875	21.8000	9.75815	235.00	751.40	1.4202	42.30	40.29	158.0	158.4	-0.4
128	2,2,5,5-tetramethylhexane	2,2,5,5-MMMM-C6	9,14,7,6,9	10	121	3.5656	<b>449</b>	14.8369	6.9651	21.5500	9.73963	229.80	714.80	1.4055	21.10	39.37	137.2	153.2	-16.0
129	2,3,4,5-tetramethylhexane	2,3,4,5-MMMM-C6	9,12,12,8,4	10	121	3.8147	449	14.7566	6.9938	22.1000	9.73832	243.10	745.60	1.4204	42.70	40.71	155.5	158.4	-2.9
130	2,3,3,5-tetramethylhexane	2,3,3,5-MMMM-C6	9,13,11,8,4	10	120	3.8745	451	14.7506	7.0154	21.4667	9.73562	234.00	744.90	1.4196	41.20	39.16	152.9	159.7	-6.8
131	2,4-dimethyl-3-ethylhexane	2,4-MM-3-E-C6	9,11,12,10,3	10	122	3.8007	455	14.6750	6.9499	22.1000	9.73353						164.0	157.6	6.4
132	3,3-diethylhexane	3,3-EE-C6	9,11,13,9,3	10	121	3.8665	461	14.6542	6.9565	22.4333	9.72676	242.50	757.50	1.4235	51.				

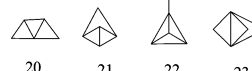
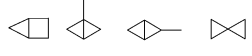
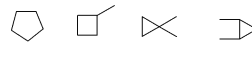
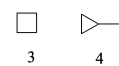
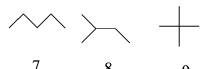


**Figure 1.** The minimal identity tree (1), a cyclic graph (2), and the minimal identity graph (3). Vertices  $v_i$  are in black and edges  $e_i$  are in red.

#### 4-Vertex graphs



#### 5-Vertex graphs



27

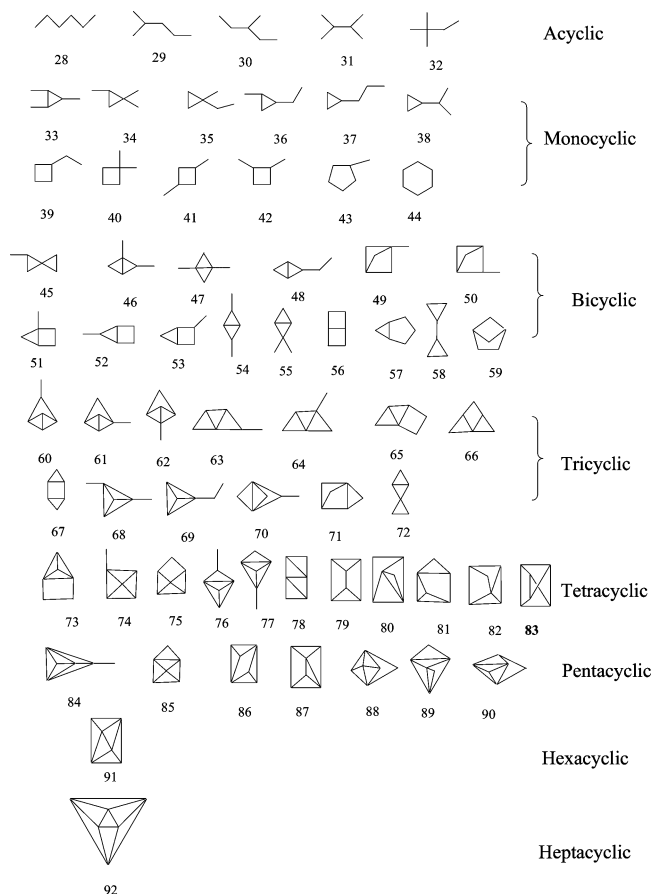


**Figure 2.** All possible chemical graphs with 4 and 5 vertices plus a nonplanar graph, 27.

and compress them into a single number, instead of looking at the second and third numbers ( $p_2$  and  $p_3$ ) in the sequence. For this purpose we follow an example that was introduced earlier for one of the first topological indices based on a sequence of numbers (namely, vertex degrees in this case) that sum up to the same total for isomers, namely the index  $M_1 = \sum_i v_i^2$ , one of the Zagreb indices.<sup>5</sup> Randić reported similar investigations compressing path numbers into one number, by summing powers of path numbers  $\sum_i p_i^k$  (with  $k$  assuming six negative values between  $-3$  and  $-0.5$  and four positive values between  $0.5$  and  $2$ ) but restricted the set of graphs to octane isomers.<sup>22</sup> He called such indices “variable Wiener numbers”. In the present paper, we examine all alkanes with 3–12 carbon atoms as well as all chemical graphs (cyclic and acyclic) with 4–6 carbon atoms, and we explore correlations with a few properties of the corresponding hydrocarbons.

#### DEFINITIONS

A *graph* is a set  $V$  of elements ( $v_i$ ) called vertices and a set  $E$  of elements ( $e_i$ ) called edges, that couple two distinct vertices. The number of edges meeting at a vertex is called the *degree* of that vertex. The sum of all vertex degrees of



**Figure 3.** All possible chemical graphs with 6 vertices, plus a nonplanar graph 83 ( $K_{3,3}$ ).

a graph is twice the number of edges. A *walk* in a graph is an alternating sequence of vertices and edges, beginning and ending with vertices, in which each edge is incident with the two vertices immediately preceding and following it; an example of a walk is the following:  $v_1e_1v_2e_2v_3e_3v_4e_3v_3$ , as presented in Figure 1 for graph 1. A *trail* is a walk whose edges are all distinct; an example for graph 2 in Figure 1 is the following:  $v_1e_5v_5e_7v_3e_2v_2e_6v_5e_8v_4$ . A *path* is a walk or a trail whose vertices (and thus necessarily all edges) are distinct; an example for graph 2 in Figure 1 is the following:  $v_1e_5v_5e_7v_3e_2v_2$ . In a *connected graph* any two vertices are connected by one or more paths. All graphs discussed in this paper will be connected graphs. A *cycle* (or a closed walk) is a walk with  $\geq 3$  distinct edges where the initial and final vertices are identical;  $v_1e_1v_2e_2v_3e_3v_4e_1$  in graph 2 is a cycle. A *tree* is a connected graph devoid of circuits; if  $n$  is its number of vertices (also called the order of the graph) the number of edges is  $e = n - 1$ . An *identity graph* or *identity tree* has the identity map on itself as the only automorphism. The smallest identity tree has 7 vertices (shown as 1), and the smallest identity graph (3) has 6 vertices. The *cyclomatic number*  $\mu$  of a graph is the smallest number of edges that have to be deleted in order to convert it into a tree; it can be shown that  $\mu = e - n + 1$ . For any graph, the number of edges is  $p_1 = n + \mu - 1$ .

In a tree every two vertices are joined by a unique path, identical to the unique trail with the same property. The numbers of trails and paths in a graph or a tree are limited, but the number of walks is not.



**Table 2.** Descriptors of All Possible Chemical Graphs with 4, 5, and 6 Vertices Plus the Two Kuratowski Nonplanar Graphs ( $K_5 = 27$ ,  $K_{3,3} = 83$ ) with Numberings as in Figures 2 and 3

graph	molecular path code	$n$	$\mu$	$W$	$J$	$Q$	$S$	$D$	$A$	$P$
1	3,2,1	4	0	10	1.9747	14.00	4.1464	2.7704	4.3333	3.3094
2	3,3	4	0	9	2.3237	18.00	3.4643	2.6001	4.5000	2.9567
3	4,4,4	4	1	8	2.0000	24.00	3.0007	1.8338	3.6666	3.2307
4	4,5,2	4	1	8	2.1711	22.50	2.8255	1.7940	3.5833	3.1095
5	5,8,6	4	2	7	2.4800	41.66	2.5047	1.4887	3.6666	3.2621
6	6,12,12	4	3	6	3.0000	81.00	2.3442	1.3345	4.0000	3.4494
7	4,3,2,1	5	0	20	2.1906	30.00	6.1527	3.5901	6.4166	4.5412
8	4,4,2	5	0	18	2.5395	36.00	5.4122	3.4700	6.6666	4.2307
9	4,6	5	0	16	3.0237	52.00	4.4547	3.2301	7.0000	3.7320
10	5,5,5,5	5	1	15	2.0833	50.00	4.4724	2.3292	5.2083	4.4026
11	5,6,6,2	5	1	16	2.0796	50.50	4.2745	2.3155	5.2500	4.3058
12	5,8,4	5	1	16	2.2034	52.50	4.0580	2.1586	5.1666	3.8118
13	5,7,5,1	5	1	17	1.9989	50.00	4.0494	2.2778	5.2083	4.1704
14	5,6,4,2	5	2	14	2.1908	27.00	3.7872	1.4964	3.2777	3.3659
15	6,9,10,7	5	2	14	2.1989	88.66	3.7519	1.8887	5.1944	4.4568
16	6,11,10,2	5	3	14	2.3822	65.25	2.5855	1.3784	3.8322	3.6637
17	6,10,10,4	5	3	15	2.1921	63.00	2.6934	1.3966	3.8333	3.7556
18	6,10,8,4	5	2	14	2.3011	72.00	3.4807	1.8245	4.8888	4.2253
19	6,9,12,6	5	2	14	2.1875	99.00	3.7877	1.9059	5.3333	4.5007
20	7,14,17,10	5	3	13	2.4865	158.50	3.4184	1.6700	5.5416	4.6265
21	7,13,18,14	5	3	13	2.3890	184.50	3.5580	1.6991	5.7500	4.7577
22	7,15,18,6	5	3	13	2.5555	158.50	3.3029	1.6524	5.5000	4.5292
23	7,15,18,6	5	3	13	2.5818	158.50	3.3029	1.6527	5.5000	4.5292
24	8,19,28,18	5	4	12	2.8044	306.60	3.3444	1.5666	6.2666	4.9582
25	8,18,28,24	5	4	12	2.7111	349.60	3.4525	1.5875	6.4666	5.0682
26	9,24,42,36	5	5	11	3.1374	619.50	3.3967	1.5185	7.3333	5.3912
27	10,30,60,60	5	6	10	3.5740	1171.42	3.4473	1.4883	8.5714	5.8132
28	5,4,3,2,1	6	0	35	2.3390	55.00	8.3917	4.3602	8.7000	5.8046
29	5,5,3,2	6	0	32	2.6272	63.00	7.6212	4.2703	9.0000	5.5243
30	5,5,4,1	6	0	31	2.7541	67.00	7.4742	4.2607	9.0833	5.4719
31	5,6,4	6	0	29	2.9935	77.00	6.6893	4.1208	9.3333	5.1228
32	5,7,3	6	0	28	3.1684	83.00	6.6147	4.1308	9.5000	5.1068
33	6,9,9,3	6	1	27	2.3112	103.50	5.0904	2.6910	7.1250	5.0691
34	6,10,8,2	6	1	26	2.4134	102.00	4.9270	2.6634	7.0833	4.9678
35	6,9,7,2	6	1	27	2.2764	85.00	4.7541	2.5925	6.6666	4.8121
36	6,8,7,4,1	6	1	29	2.0940	83.00	5.4615	2.7223	6.7666	5.2497
37	6,7,5,4,2	6	1	31	1.8764	65.00	5.3728	2.6504	6.2833	5.1221
38	6,8,6,4	6	1	28	2.1289	76.00	4.8639	2.5908	6.5000	4.8533
39	6,7,8,4,2	6	1	29	2.0142	84.50	5.6684	2.7491	6.7833	5.3639
40	6,9,8,4	6	1	26	2.3110	98.50	5.1384	2.6966	7.0833	5.0938
41	6,8,8,6	6	1	28	2.1120	100.00	5.2777	2.7093	7.0833	5.1669
42	6,8,9,4,1	6	1	27	2.2022	99.00	5.6381	2.7810	7.1000	5.3943
43	6,7,7,7,2	6	1	26	2.1847	93.50	5.9001	2.7998	6.9916	5.5176
44	6,6,6,6,6	6	1	27	2.0000	90.00	6.1232	2.7968	6.8500	5.5974
45	7,12,12,8,2	6	2	25	2.1694	135.00	4.6057	2.1743	6.4666	5.2780
46	7,13,15,7,1	6	2	25	2.2325	164.33	4.5904	2.2000	6.8166	5.3124
47	7,14,15,6	6	2	24	2.3389	168.66	4.2365	2.1394	6.8333	5.0531
48	7,11,12,8,4	6	2	28	1.9226	131.33	4.7517	2.1881	6.4333	5.3691
49	7,12,15,12	6	2	24	2.1874	187.33	4.4826	2.1785	7.0000	5.2327
50	7,11,16,10,4	6	2	25	2.1087	180.66	5.0416	2.2753	7.0444	5.6441
51	7,12,14,10,2	6	2	24	2.2160	164.33	4.8094	2.2324	6.8555	5.4669
52	7,11,14,11,4	6	2	25	2.1133	167.66	5.0062	2.2600	6.9055	5.6025
53	7,11,13,12,3	6	2	25	2.0838	164.00	4.9210	2.2397	6.8111	5.5305
54	7,12,14,10,2	6	2	27	2.0573	164.33	4.8094	2.2321	6.8555	5.4669
55	7,13,14,8	6	2	25	2.2144	159.33	4.2731	2.1345	6.7222	5.0632
56	7,10,14,10,8	6	2	25	2.0280	169.66	5.1805	2.2764	6.9222	5.7089
57	7,10,11,12,8	6	2	24	2.0954	159.33	5.1396	2.2541	6.7555	5.6543
58	7,10,8,8,4	6	2	27	1.9166	97.66	4.4885	2.0926	5.8222	5.0942
59	7,10,13,14,6	6	2	23	2.1624	183.33	5.2012	2.2844	7.0111	5.7329
60	8,15,22,22,8	6	3	24	2.1180	330.25	4.7277	2.0162	7.4833	5.9425
61	8,16,23,22,4	6	3	23	2.2347	337.25	4.5789	1.9992	7.4916	5.8326
62	8,16,24,20,6	6	3	23	2.2599	333.00	4.6629	2.0171	7.5500	5.9083
63	8,16,22,18,6	6	3	24	2.1950	291.00	4.5525	1.9851	7.3416	5.7908
64	8,17,23,17,3	6	3	23	2.3104	295.00	4.4004	1.9665	7.2541	5.6744
65	8,15,21,18,11	6	3	23	2.2033	293.75	4.7102	2.0040	7.3000	5.9086
66	8,15,20,21,10	6	3	22	2.2644	307.50	4.7297	2.0085	7.3541	5.9272
67	8,14,18,22,12	6	3	23	2.1517	303.00	4.7416	1.9948	7.2250	5.9090
68	8,18,25,14,2	6	3	23	2.3700	303.25	4.3064	1.9585	7.3083	5.6092
69	8,16,21,12,6	6	3	25	2.1090	235.25	4.3311	1.9274	6.8000	5.5650
70	8,17,24,14,4	6	3	24	2.2677	285.25	4.3983	1.9641	7.2000	5.6687
71	8,15,22,22,8	6	3	22	2.2621	330.25	4.7274	2.0163	7.4833	5.9425
72	8,15,18,16,8	6	3	24	2.1484	233.25	4.4437	1.9365	6.7750	5.6407
73	9,20,32,30,20	6	4	21	2.4132	561.00	4.6150	1.8776	8.2333	6.3356

Table 2. (Continued)

graph	molecular path code	$n$	$\mu$	$W$	$J$	$Q$	$S$	$D$	$A$	$P$
74	9,21,35,36,10	6	4	22	2.3844	628.60	4.5325	1.8798	8.4333	6.2924
75	9,19,31,37,26	6	4	21	2.3633	689.60	4.8216	1.9150	8.6566	6.5375
76	9,21,34,30,12	6	4	23	2.3254	544.40	4.4705	1.8590	8.1466	6.0867
77	9,22,36,30,8	6	4	22	2.4561	565.00	4.3992	1.8567	8.2200	6.1645
78	9,20,31,32,17	6	4	22	2.3540	551.00	4.5633	1.8665	8.1466	6.2829
79	9,18,30,42,30	6	4	21	2.3142	793.80	4.9350	1.9322	8.9000	6.6420
80	9,20,34,32,20	6	4	22	2.3543	612.20	4.6863	1.8971	8.4666	6.4207
81	9,19,30,38,22	6	4	21	2.3622	654.00	4.7381	1.8964	8.4800	6.4507
82	9,19,33,38,24	6	4	21	2.3624	710.20	4.8332	1.9237	8.7600	6.5614
83	9,18,36,36,36	6	4	31	2.2900	858.60	5.0480	1.9642	9.2400	6.7741
84	10,27,51,54,18	6	5	21	2.6042	1111.66	4.5155	1.8044	9.6000	6.7488
85	10,25,46,59,34	6	5	20	2.5811	1246.33	4.7427	1.8341	9.8972	6.9654
86	10,24,44,64,40	6	5	20	2.5340	1384.66	4.8364	1.8473	10.1111	7.0563
87	10,24,45,62,45	6	5	20	2.5355	1428.33	4.8917	1.8590	10.2500	7.1183
88	10,25,48,62,32	6	5	20	2.5800	1316.16	4.7701	1.8452	10.0666	7.0074
89	10,26,48,52,28	6	5	21	2.5744	1094.66	4.6150	1.8132	9.6000	6.8339
90	10,25,47,56,38	6	5	20	2.5824	1252.33	4.7775	1.8415	9.9611	7.0032
91	11,30,63,96,70	6	6	19	2.7646	2729.42	4.9855	1.8314	12.1428	7.7153
92	12,36,84,144,120	6	7	18	3.0000	5454.00	5.1978	4.8693	14.7500	8.4489

**Chemical graphs** have vertices of degree at most 4. In addition, they are graph-theoretically planar, i.e., they can be embedded on a plane without crossing edges. Kuratowski proved that all nonplanar graphs contain either the complete graph  $K_5$  or the bipartite  $K_{3,3}$  graph as subgraphs. These nonplanar graphs are included in Figures 2 and 3 (although they are nonplanar) but will not be taken into account in any correlation involving chemical graphs. Whereas graph  $K_5$  is the only nonplanar graph with 5 vertices, there are other nonplanar graphs with 6 vertices in addition to  $K_{3,3}$ .

When nonisomorphic chemical graphs have the same numerical descriptor, this descriptor is said to be degenerate; three different kinds of degeneracy will be defined further below. High degeneracy is undesirable because if one wants to correlate various properties of chemical graphs with their different structures, the descriptors should be able to differentiate sufficiently these structures so that their numerical characterization should be able to reflect their similarity or dissimilarity.

#### NOTATION

The degree of a vertex  $i$  is denoted by  $v_i$ . Let  $p_i$  and  $t_i$  denote the number of paths (or trails, respectively) of length  $i$  ( $p_1 \equiv t_1$  denotes the number of edges in the graph). For trees,  $p_i \equiv t_i$ . For linear  $n$ -alkanes, the molecular path code  $p_1, p_2, \dots, p_{n-1} \equiv t_1, t_2, \dots, t_{n-1}$  is identical to the sequence of natural numbers starting with 1 and ending with  $n - 1$ .

Let  $\Sigma_i p_i$  denote the sum of all paths in a graph;  $\Sigma_i p_i^2$  will denote the sums of squares of all paths. For any tree with  $n$  vertices such as an alkane with  $n$  carbon atoms,  $\Sigma_i p_i = n(n - 1)/2$ . In a tree, the parities of  $\Sigma_i p_i$  and  $\Sigma_i p_i^2$  are interesting: both are even for  $n \equiv 0(\text{mod } 4)$  and  $n \equiv 1(\text{mod } 4)$ ; they are odd for  $n \equiv 2(\text{mod } 4)$  and  $n \equiv 3(\text{mod } 4)$ .

#### INDEX $Q$ : STOCHASTIC, APPARENT, AND REAL DEGENERACY, NORMAL ORDERING OF ALKANES

We investigated a new *Quadratic molecular descriptor*  $Q = \Sigma_i p_i^2 / (\mu_i + 1)$ . For trees, because  $\mu = 0$ , we have  $Q = \Sigma_i p_i^2$ . A linear  $n$ -alkane has  $Q = n(n - 1)(2n - 1)/6$ . All the above equalities may be easily checked by examining Table 1.

For the lower alkanes with 3–10 carbon atoms, Table 1 presents molecular path codes, and  $Q$  values. It will be observed that all molecular path codes (path sequences) are different for alkanes with up to 8 carbon atoms. Table 2 presents cyclomatic numbers  $\mu$ , molecular path codes, and  $Q$  values for all 92 possible acyclic and (poly)cyclic chemical graphs with 4–6 vertices devoid of multiple bonds that are displayed in Figures 2 and 3, plus the two Kuratowski nonplanar graphs **27** and **83**.

**Properties of Index  $Q$ .** It can be seen from Table 1 that the new index  $Q$  increases with the graph size, but the variation ranges overlap so that  $Q$  for the most branched heptane isomer is larger than  $Q$  for the two least branched isomers of octane. Following the discussion by Bonchev,<sup>23</sup> within an isomeric series (we will use octanes as examples),  $Q$  increases with (i) the number of branches (M-C7 < MM-C6 < MMM-C5); (ii) the centrality of branches (2M-C7 < 3M-C7 < 4M-C7; 22MM-C6 < 33MM-C6); (iii) the length of branches (3M-C7 < 3E-C6); and (iv) the clustering or multiplicity of branches (24MM-C6 < 22MM-C6). Sometimes, such factors counterbalance each other, leading to *accidental (stochastic) degeneracy* because sums of squares of integers may coincidentally afford the same integer (e.g., for two pairs of heptane isomers,  $Q = 125$  for 22MM-C5 and 23MM-C5;  $Q = 117$  for 24MM-C5 and 3E-C5, or for one pair of octane isomers  $Q = 210$  for 224MMM-C5 and 234MMM-C5).

The influence exerted by the cyclicity of graphs will be exemplified using all possible cyclic and polycyclic graphs with 4–6 vertices, illustrated by Figures 2 and 3. Stereochemistry (diastereoisomerism) is ignored, because the differences between properties of *E* and *Z* stereoisomers are minor (e.g., for the normal boiling points, NBP, only a few

	$W$	$J$	$Q$	$S$	$D$	$A$	$P$
$W$	1	0.3746	0.7848	0.9804	0.9533	0.9027	0.9776
$J$	0.3746	1	0.8168	0.3083	0.5617	0.7234	0.4433
$Q$	0.7848	0.8168	1	0.7032	0.8449	0.9490	0.7856
$S$	0.9804	0.3083	0.7032	1	0.9568	0.8657	0.9860
$D$	0.9533	0.5617	0.8449	0.9568	1	0.9649	0.9889
$A$	0.9027	0.7234	0.9490	0.8657	0.9649	1	0.9276
$P$	0.9776	0.4433	0.7856	0.9860	0.9889	0.9276	1

**Figure 4.** Interrelation ( $R^2$  values) of TIs for alkanes with 3–10 carbon atoms. Italics denote high intercorrelations with  $R^2 > 0.97$ .

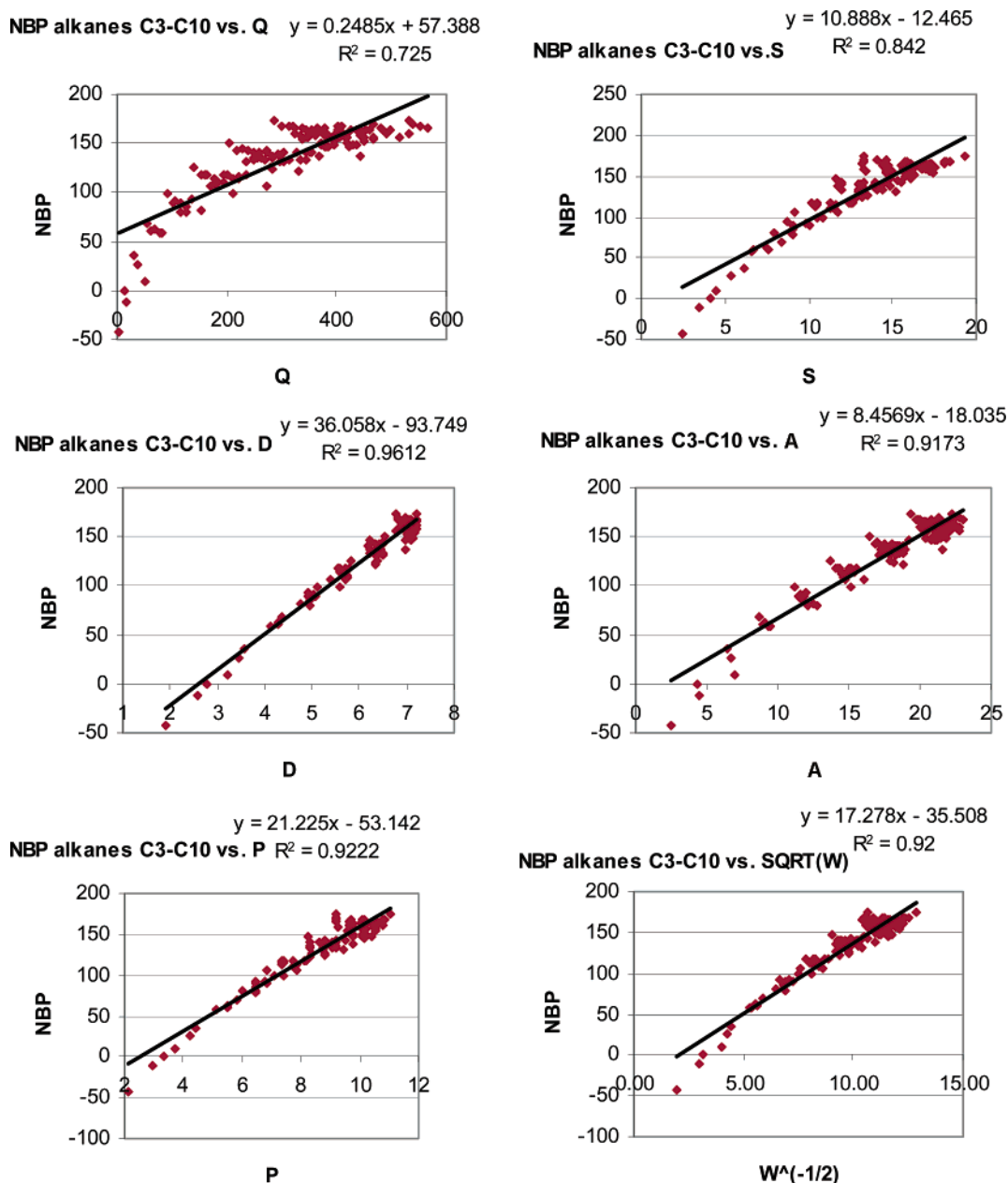


Figure 5. Plots of normal boiling points of C<sub>3</sub>–C<sub>10</sub> alkanes versus TIs.

degrees Celsius). In monocyclic graphs (for a ring with  $n$  carbon atoms without side chains, the path sequence is  $1, 2, 3, \dots, n-1$ ), index  $Q$  increases with the size of the ring. However, for cyclic graphs, isomers no longer have the same  $\sum p_i$  value, as they had in the case of trees. On comparing isomeric graphs with the same cyclomatic number, it can be seen that  $Q$  increases with (i) the number of branches (propylcyclopropane < 1-ethyl-2-methylcyclopropane < both trimethylcyclopropanes; propylcyclopropane < isopropylcyclopropane); (ii) the centrality of branches (1,1,2-trimethylcyclopropane < 1,2,3-trimethylcyclopropane); and (iii) the ring size (among all bicyclic C<sub>6</sub>H<sub>10</sub> isomers, bis-cyclopropane has the smallest  $Q$  value and bicyclo[2.1.1]hexane the largest one). However, the length, clustering, or multiplicity of branches cannot vary independently of the cyclomatic number, and the resulting dependence of the molecular path code and  $Q$  index is rather complicated.

**Degeneracies of molecular path codes (path sequences)** start with the sequence 8,10,10,6,2 shared by two nonane isomers having  $Q = 304$ . The degeneracy of molecular path codes automatically leads to the degeneracy of  $Q$  values, but the reverse is not true, as will be shown in the next paragraph.

**Degeneracies of  $Q$  values** start, as indicated earlier, with two pairs of heptane isomers ( $Q = 117$  and  $125$ ) and continue with one pair of octanes ( $Q = 210$ ) as well as three pairs of nonanes ( $Q = 248$ ,  $344$ , and  $304$ ). Among all of these, only the last one is a *real degeneracy* caused by the *same molecular path code, including the order of path numbers*; the remainder are *apparent degeneracies* (i.e. the same numbers in the sequence but in a different order as in the case of  $Q = 344$ ) or *stochastic (chance degeneracies, i.e. coincidence of sums derived from different sets of numbers as in the case of  $Q = 248$ )*. For decane isomers, there are

**Table 3.** Alkanes and (Poly)cycloalkanes with Known NBP's having 4–6 Carbon Atoms with their Data and NBP's, Sorted by  $n$  and then by  $D^a$ 

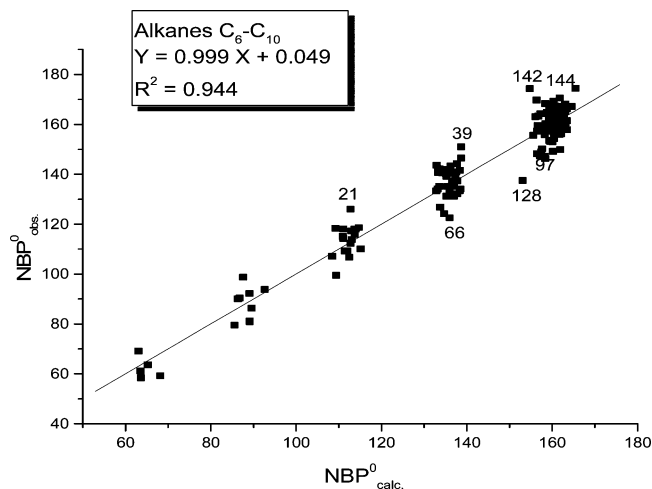
no.	alkanes C <sub>4</sub> –C <sub>6</sub> and (poly)cycloalkanes C <sub>4</sub> –C <sub>6</sub>	molecular path code	$n$	$\mu$	$W$	$J$	$Q$	$S$	$D$	$A$	$P$	NBP (°C) (exp)	NBP (°C) (calc)	resid
1	butane	3,2,1	4	0	10	1.9747	14.00	4.1462	2.7724	4.3333	3.3094	–0.1	5.1	–5.2
2	2-methylpropane	3,3	4	0	9	2.3237	18.00	3.4641	2.5980	4.5000	2.9568	–11.2	–4.5	–6.7
3	cyclobutane	4,4,4	4	1	8	2.0000	24.00	3.0000	1.8333	3.6666	3.2307	12.6	8.8	3.8
4	methylcyclopropane	4,5,2	4	1	8	2.1711	22.50	2.8251	1.7940	3.5833	3.1095	0.7	5.3	–4.6
5	bicyclo[1.1.0]butane	5,8,6	4	2	7	2.4800	41.66	2.5046	1.4887	3.6666	3.2621	4.3	11.9	–7.6
7	pentane	4,3,2,1	5	0	20	2.1906	30.00	6.1527	3.5901	6.4166	4.5412	36.1	37.5	–1.4
8	2-methylbutane	4,4,2	5	0	18	2.5395	36.00	5.4122	3.4700	6.6666	4.2307	27.0	28.8	–1.8
9	2,2-dimethylpropane	4,6	5	0	16	3.0237	52.00	4.4547	3.2301	7.0000	3.7320	9.5	15.1	–5.6
10	cyclopentane	5,5,5,5	5	1	15	2.0833	50.00	4.4724	2.3292	5.2083	4.4026	49.3	41.4	7.9
11	methylcyclobutane	5,6,6,2	5	1	16	2.0796	50.50	4.2745	2.3155	5.2500	4.3058	36.3	38.5	–2.2
13	e-1,2-dimethylcyclopropane	5,7,5,1	5	1	17	1.9989	50.00	4.0494	2.2778	5.2083	4.1704	33.0	34.6	–1.6
12	1,1-dimethylcyclopropane	5,8,4	5	1	16	2.2034	52.50	4.0580	2.1586	5.1666	3.8118	20.6	24.4	–3.8
19	bicyclo[1.1.1]pentane	6,9,12,6	5	2	14	2.1875	99.00	3.7877	1.9059	5.3333	4.5007	46.0	47.1	–1.1
15	bicyclo[2.1.0]cyclopentane	6,9,10,7	5	2	14	2.1989	88.66	3.7519	1.8887	5.1944	4.4568	45.5	45.9	–0.4
18	spiro[2.2]pentane	6,10,8,4	5	2	14	2.3011	72.00	3.4807	1.8245	4.8888	4.2253	39.0	39.3	–0.3
14	ethylcyclopropane	5,6,4,2	5	1	14	1.9994	27.00	2.6999	1.4964	3.2777	3.3654	35.9	15.1	20.8
28	hexane	5,4,3,2,1	6	0	35	2.3390	55.00	8.3917	4.3602	8.7000	5.8046	68.8	71.1	–2.3
29	2-methylpentane	5,5,3,2	6	0	32	2.6272	63.00	7.6212	4.2703	9.0000	5.5243	60.9	63.1	–2.2
30	3-methylpentane	5,5,4,1	6	0	31	2.7541	67.00	7.4742	4.2607	9.0833	5.4719	63.3	61.6	1.7
32	2,3-dimethylbutane	5,7,3	6	0	28	3.1684	83.00	6.6147	4.1308	9.5000	5.1068	58.9	51.3	7.6
31	2,2-dimethylbutane	5,6,4	6	0	29	2.9935	77.00	6.6893	4.1208	9.3333	5.1228	58.1	51.8	6.3
43	methylcyclopentane	6,7,7,7,2	6	1	26	2.1847	93.50	5.9001	2.7998	6.9916	5.5176	71.9	72.5	–0.6
44	cyclohexane	6,6,6,6,6	6	2	27	2.0000	90.00	6.1232	2.7968	6.8500	5.5974	80.8	74.9	5.9
42	1,2-dimethylcyclobutane	6,8,9,4,1	6	1	27	2.2022	99.00	5.6381	2.7810	7.1000	5.3943	59.0	68.8	–9.8
39	ethylcyclobutane	6,7,8,4,2	6	1	29	2.0142	84.50	5.6684	2.7491	6.7833	5.3639	70.8	68.1	2.7
41	1,3-dimethylcyclobutane	6,8,8,6	6	1	28	2.1120	100.00	5.2777	2.7093	7.0833	5.1669	57.5	62.3	–4.8
33	1,2,3-trimethylcyclopropane	6,9,9,3	6	1	27	2.3112	103.50	5.0904	2.6910	7.1250	5.0691	63.0	59.4	3.6
34	1,1,2-trimethylcyclopropane	6,10,8,2	6	1	26	2.4134	102.00	4.9270	2.6634	7.0833	4.9678	54.0	56.5	–2.5
37	propylcyclopropane	6,7,5,4,2	6	1	31	1.8764	65.00	5.3728	2.6504	6.2833	5.1221	69.1	61.3	7.8
35	1-ethyl-2-methylcyclopropane	6,9,7,2	6	1	27	2.2764	85.00	4.7541	2.5925	6.6666	4.8121	59.0	52.2	6.8
38	isopropylcyclopropane	6,8,6,4	6	1	28	2.1289	76.00	4.8639	2.5908	6.5000	4.8533	58.3	53.5	4.8
59	bicyclo[2.1.1]hexane	7,10,13,14,6	6	2	23	2.1624	183.33	5.2012	2.2844	7.0111	5.7329	71.0	82.4	–11.4
56	bicyclo[2.2.0]hexane	7,10,14,10,8	6	2	25	2.0280	169.66	5.1805	2.2764	6.9222	5.7089	80.0	81.7	–1.7
57	bicyclo[3.1.0]hexane	7,10,11,12,8	6	2	24	2.0954	159.33	5.1396	2.2541	6.7555	5.6543	81.0	80.2	0.8
51	1-methylbicyclo[2.1.0]pentane	7,12,14,10,2	6	2	24	2.2160	164.33	4.8094	2.2324	6.8555	5.4669	80.0	74.6	5.4
54	2,4-dimethylbicyclo[1.1.0]butane	7,12,14,10,2	6	2	27	2.0573	164.33	4.8094	2.2321	6.8555	5.4669	54.5	74.6	–20.1
58	bis-cyclopropane	7,10,8,8,4	6	2	27	1.9166	97.66	4.4885	2.0926	5.8222	5.0942	76.2	64.1	12.1

<sup>a</sup> Numberings are as in Figures 2 and 3.

more cases of stochastic degeneracies:  $Q = 337, 343, 361, 371, 379, 383, 395, 419, 435$ , and  $443$ ; in the last case there are three different molecular path codes yielding the same  $Q = 443$  value as well as a really degenerate path sequence. For  $Q = 467$  there is an apparent and a stochastic degeneracy.

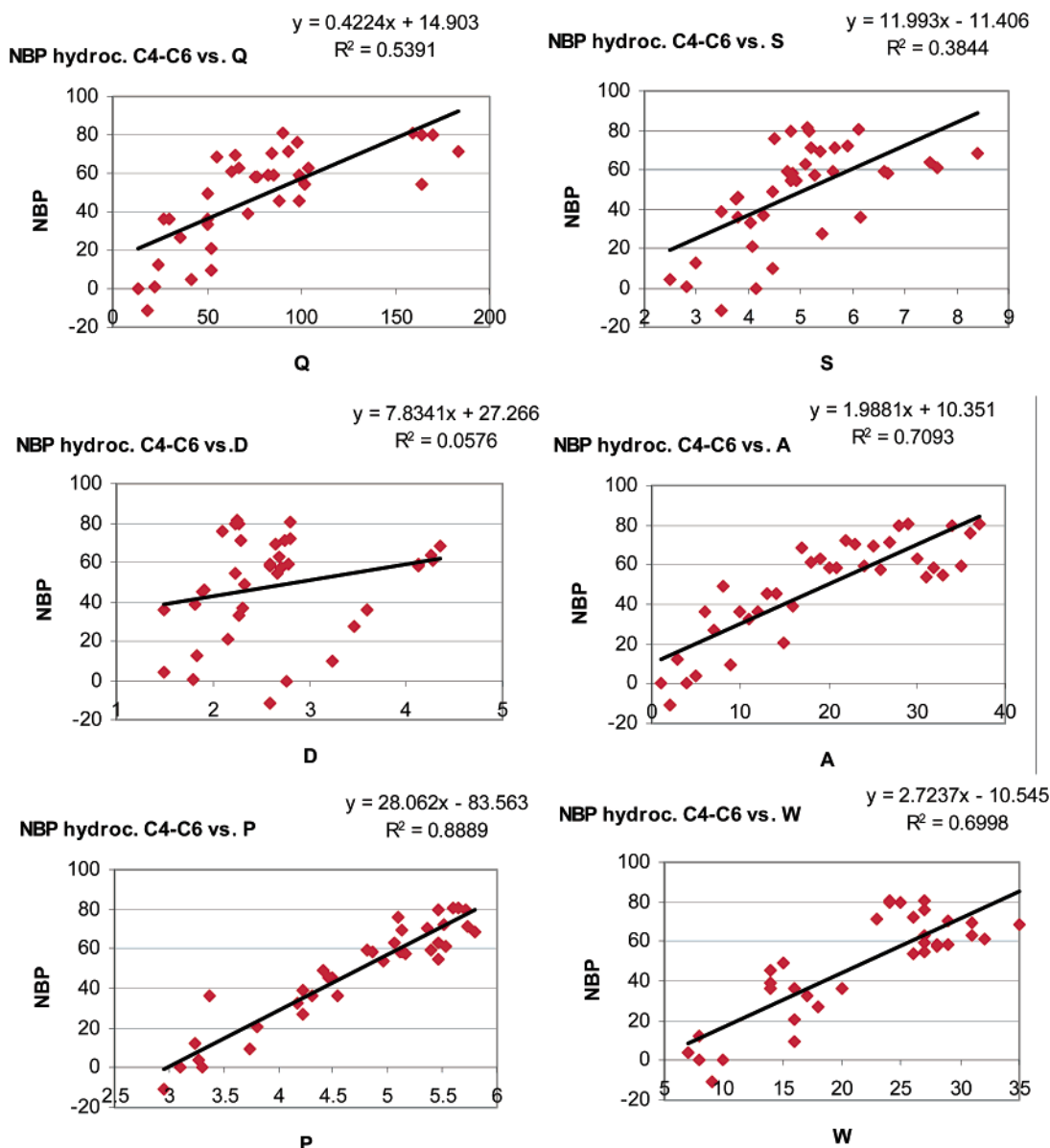
As  $n$  increases, the degeneracies of  $Q$  and of molecular path codes increase: there are 3  $Q$  values for nonanes, 9 for decanes, 37 for undecanes, and 90 values for dodecanes which are shared by more than one alkane (these are real, apparent, and stochastic degeneracies). One has to take into account that for any set of alkane isomers, the number of graphical partitions of  $\sum ip_i = n(n-1)/2$  must increase less rapidly than the number of isomers of such alkanes which are 4-trees (the vertex degree is a number equal to or smaller than 4). As a consequence, the number of degeneracies will rise with increasing  $n$  values.

It should be noted that the  $Q$  index is somewhat less degenerate than other first-generation topological indices expressed by integer numbers which do not run into large numbers such as Bertz's graph derivatives (see the next section). The Wiener index<sup>1</sup>  $W$  for alkanes can be expressed with the aid of the molecular path code as  $W = \sum ip_i$ . Indeed, indices  $Q$  and  $W$  are degenerate for the same two pairs of heptane isomers; Bonchev's overall Wiener's index<sup>23</sup> ( $OW$ ) is degenerate only for one of these pairs (3-ethylpentane and

**Figure 6.** Plot of observed vs calculated values for NBP of C<sub>6</sub>–C<sub>10</sub> alkanes according to the biparametric eq 1 (see Table 1).

2,4-dimethylpentane). For octane isomers,  $W$  is degenerate for two pairs of isomers,  $Q$  for one pair, and  $OW$  for none. However,  $W$  becomes extremely degenerate for higher alkanes: as an example, out of 35 decane constitutional isomers, there are 29 shared  $W$  values (4 pairs, 3 triples, and 3 quadruples), whereas for  $Q$  the number of shared





**Figure 7.** Plots of monoparametric correlations of NBP for cyclic and acyclic saturated hydrocarbons C<sub>4</sub>–C<sub>6</sub> vs TIs.

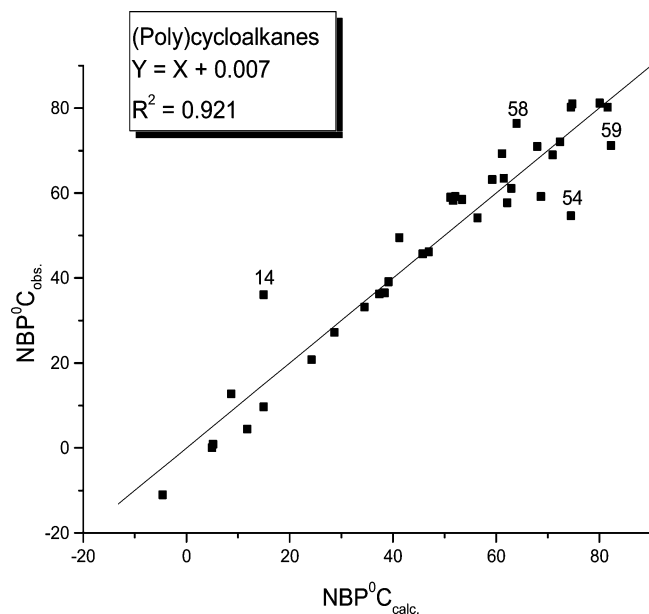
values for decane isomers is only 12. For cyclic graphs,  $Q$  is no longer an integer and starts to be degenerate for the pair ethylcyclopropane and 1,1-dimethylcyclopropane which have  $Q = 344$ .

#### INDEX S: APPARENT AND REAL DEGENERACY, NORMAL ORDERING OF ALKANES

Taking into account the fairly high degeneracy (due to the fact that it is an integer with fairly low values) and other drawbacks of the quadratic index  $Q$  (derived from the fact that squaring numbers of path numbers with various lengths places more weight on paths between remote vertices with little interaction among them) we investigated ways to stress the contribution of closely spaced vertices. The sum of reciprocal values of numbers in the molecular path code did not produce satisfactory correlations and was abandoned. On the other hand, we explored in detail a different index,  $S$ , which is second-generation even for alkanes. It is based on summing the ratio between square-roots of the numbers in the path sequence and the cyclomatic number plus one:  $S = \sum p_i^{1/2}/(\mu_i + 1)$ . For a comparison between the variation

of normal boiling points (NBPs) for alkanes with 3–10 carbon atoms, one will see in a later section from Figure 5 that, although in all cases the variation is not linear, the correlation is considerably better for index  $S$  than for index  $Q$ . This is due mostly to the fact that  $Q$  increases both with size and branching, whereas  $S$  increases with size but decreases with branching just as NBPs for all alkanes do. Index  $S$  also decreases with cyclicity.

Index  $S$  is not an integer but a real number, and all its degeneracies result from matching the numbers in the molecular path code (irrespective of the order in which they appear in the sequence), i.e. there are practically no chance degeneracies. A real degeneracy for nonane isomers occurs for the pair with  $Q = 304$  and  $S = 13.016$ . In addition, the first apparent degeneracy starts with nonane isomers, as shown in the preceding section (the two isomers mentioned above with  $Q = 344$  and  $S = 11.9043$ ). There are also 2 apparent degeneracies for decane isomers (in addition to one real degeneracy), 4 apparent degeneracies for undecane isomers, and 25 apparent degeneracies for dodecane isomers (including two triplets). The next section describes a third



**Figure 8.** Plot of observed vs calculated NBP values for the biparametric eq 2.

new descriptor based on the molecular path code, which has only real degeneracies.

#### INDEX D: ONLY REAL DEGENERACY, BUT ABNORMAL ORDERING OF ALKANES

Unlike the two previous indices, which were degenerate irrespective of the order in which path numbers occurred in the molecular path code, a third index derived from index *S* (**D**istance-reduced index, *D*) is degenerate only when molecular path codes coincide both in path numbers and in their order (real degeneracies). It is obtained by dividing the square root of each path number by the corresponding distance and by the cyclomatic number plus one:  $D = \sum_i \{p_i^{1/2}/[i(\mu+1)]\}$ . Thus, index *D* is degenerate for one pair of nonane isomers and one pair of decane isomers (boldface characters in Table 1). Then there are 12 degenerate pairs and one triplet for the undecane isomers; for the dodecane isomers there are 16 degenerate pairs.

For the 37 (poly)cyclic graphs from Table 3 index *D* is the only one that orders graphs approximately according to their increasing cyclomatic number (except for ethylcyclopropane and cyclohexane), because *D* increases markedly with cyclicity and decreases markedly with branching (mimicking the same variation trend for NBP).

Although index *D* orders alkanes with 3–11 carbon atoms in approximately normal fashion with the normal alkane having an extreme value among all its isomers, for dodecane isomers this is no longer true. Moreover, even with smaller alkanes the branching order for monomethyl and dimethyl isomers is scrambled.

#### INDEX A: STOCHASTIC, APPARENT, AND REAL DEGENERACY, NORMAL ORDERING OF ALKANES

In order to explore whether a simpler formula than that corresponding to index *D* might afford a more satisfactory index, the very simple relationship was tested for a distance-Attenuated index (which will be denoted in this paper as *A*,

although for graphs that have  $\mu = 0$  the preferred notation should be *H*):  $A = \sum_i \{p_i/[i(\mu+1)]\}$ . For alkanes (acyclic tree-like graphs) this is the Harary number proposed independently in Zagreb (where it was denoted as *H*) by Nenad Trinajstić and co-workers<sup>24a</sup> and in Bucharest (where it was denoted as RDSUM).<sup>24b</sup> Index *H* was generalized recently by Estrada together with other TIs.<sup>24c</sup> Because it is related to the four new indices, we discuss it comparatively here. Although, as expected, it leads to a normal ordering of alkanes, it has a quite high degeneracy because of all possible types of degeneracy (stochastic, apparent, and real). Because only for  $\mu > 0$  this is a new index, we have not counted *A* as a new index in the title of this paper.

#### INDEX P: ONLY REAL DEGENERACY, NORMAL ORDERING OF ALKANES

Finally, we then tested the last new molecular descriptor (*P*) based on the sequence of path counts, the Path-count index:  $P = \sum_i \{p_i^{1/2}/[i^{1/2}(\mu_i+1)]\}$ . By all its properties (including the facts that for structures discussed in this paper, index *P* shares only the real degeneracies with the molecular path count and that it orders alkane isomers normally), this index appears the most promising of all five, not only by its lowest degeneracy but also because of other features. This will be proved in the next section by the satisfactory correlation with NBPs of alkanes and (poly)cycloalkanes and with other properties of alkanes.

All five TIs for undecane and dodecane constitutional isomers are presented in the Supporting Information. As indicated earlier, there is one real degeneracy of path sequences for nonane constitutional isomers. It can be seen that index *P* for  $C_{11}H_{24}$  alkanes has degeneracy for 8 pairs and one triplet and that for  $C_{12}H_{26}$  alkanes it is degenerate for 16 pairs. These degeneracies are indicated in boldface characters, and all these isomers also have degeneracies for all other indices from the Supporting Information tables except for *J*. On the other hand, index *J* has no degeneracy for any alkane with less than 12 carbon atoms, but there are 6 pairs of dodecane isomers with degenerate *J* values (3.5757, 3.7523, 3.7735, 3.9544, 4.1352, and 4.2537). Interestingly, none of these pairs has degeneracies for the five TIs, so that one may consider that *J* and *P* convey complementary information.

#### ORDERING OF ALKANE AND (POLYCYCLO)ALKANE CONSTITUTIONAL ISOMERS

The ordering of constitutional isomers of alkanes has been discussed in several previous papers, by comparing the ordering induced by several widely used topological indices.<sup>24,25</sup> Another interesting observation obtained from Table 1 is that the ordering of constitutional isomers of alkanes by indices *Q*, *S*, *D*, *R*, and *P* (up to and including heptanes) is identical with the sorting provided by the *J* index introduced earlier.<sup>26</sup> All the former five indices and index *J* are based on topological distances along the shortest paths between all graph vertices, so that there should be a certain amount of similarity between them. Of course, the accidental degeneracy of *Q* for two pairs of heptanes does not affect this ordering.

One should take into account that Bertz provided a rigorous discussion of branching in molecular graphs, as

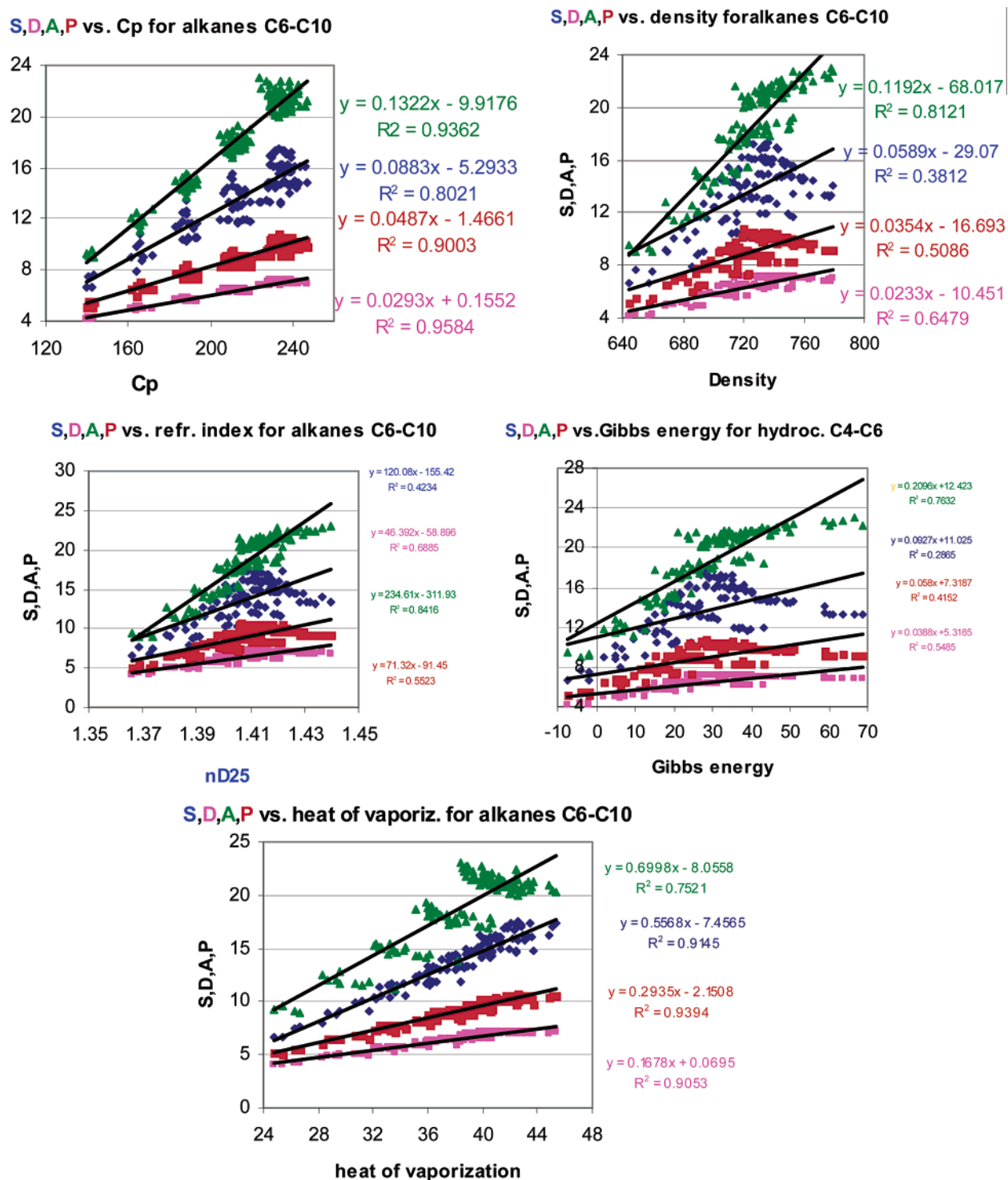


Figure 9. Plots of TIs (*S*, *D*, *A*, and *P*) versus physical properties.

reflected in the ordering of alkane isomers, arguing that graph derivatives ( $N^V$ ) provide the best logical ordering.<sup>27</sup> He wrote that

“remarkably, the ordering of the isomeric pentanes, hexanes, and heptanes using *J* is the same as the order based on the graph derivatives. Thirteen of the eighteen octanes are ordered in the same way by these two methods. It is not surprising that some quantity based on the distance matrix should parallel our order of branching, as the iterated graph derivatives incorporated information about extended connectivity as successively greater distances (cf. theorems 5

and 6). It is surprising that this particular, arbitrary formula is so successful ... Balaban has noted that “Index *J* arranges alkane isomers in an order which differs from that induced by all other topological indices.”

On the other hand, the ordering of the acyclic and (poly)-cyclic structures with 4–6 carbon atoms from Figures 2 and 3 differs substantially according to each of the five indices: indices *W*, *P*, and *D* keep the *n*-alkane as the hydrocarbon with the largest value for the TI among compounds with the same *n*, and *Q* having the lowest value, but *S* and *A* do not evidence a logical ordering.

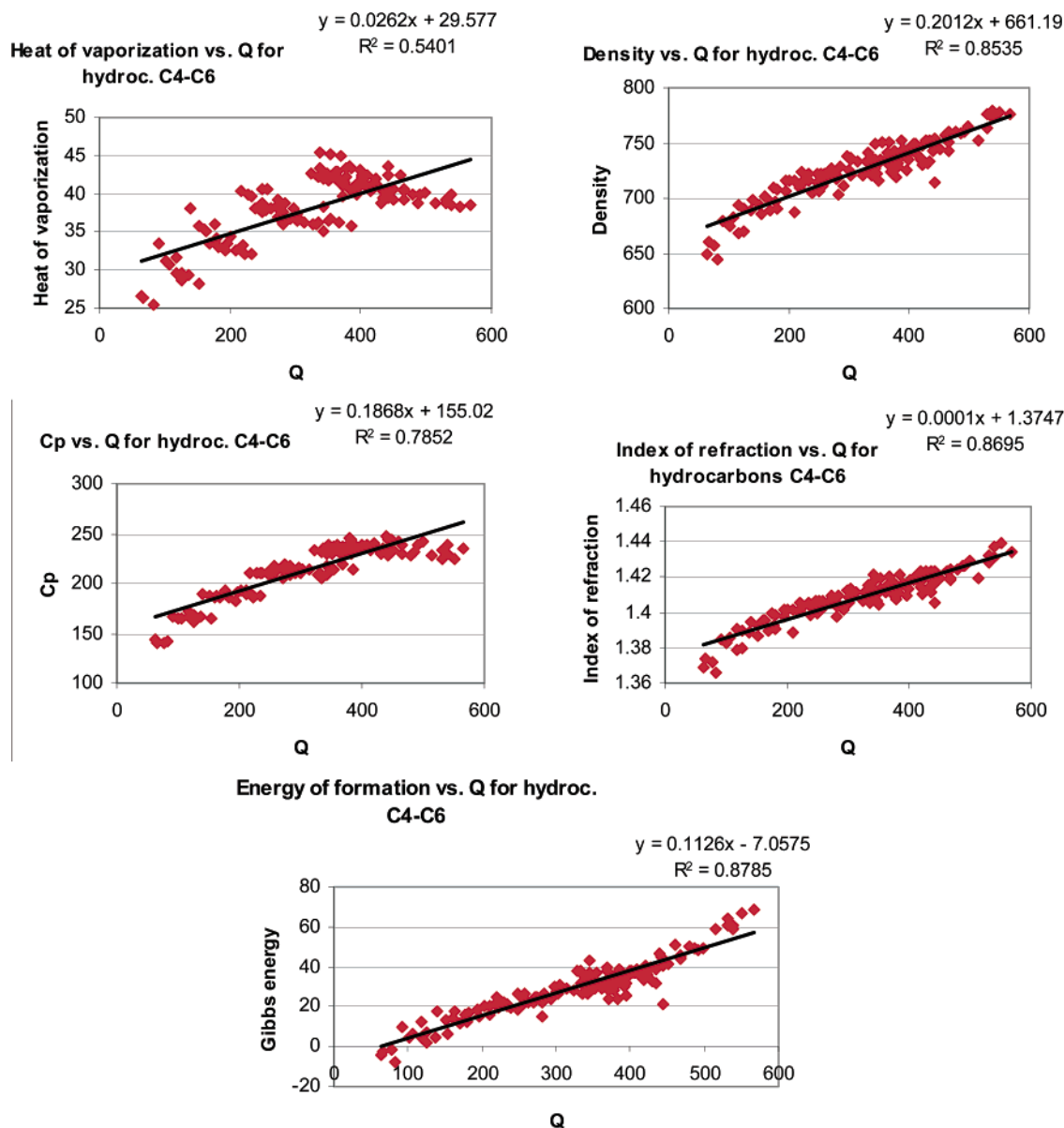


Figure 10. Plots of physical properties versus index  $Q$ .

#### QSPR STUDIES INVOLVING THE MOLECULAR DESCRIPTORS $Q$ , $S$ , $D$ , $A$ , AND $P$

On the basis of data from Table 1 (alkanes with 3–10 carbon atoms), Figure 4 presents intercorrelations between molecular descriptors discussed in the present article. Comparing these five indices among them, one finds for the set of all alkanes with 3–10 carbon atoms higher intercorrelations than  $R^2 = 0.95$  (italics in Figure 4) for all indices with square roots of path numbers, i.e., only  $Q$  is not highly correlated with the other four indices. If one includes also  $W$  and  $J$  indices, one finds that the latter is not correlated with any of the five indices, but  $W$  is highly correlated, as expected, because  $W$  also involves the sum of all distances.

By contrast, intercorrelations among all TIs based on data from Tables 2 have  $R^2$  values lower than 0.80, except for three pairs of molecular descriptors:  $(n - W)$ ,  $(S - W)$ , both with  $R^2 < 0.85$ , and  $(A - P)$  with  $R^2 = 0.918$ . Pairs of molecular descriptors with  $R^2 > 0.80$  based on data from Table 3 are  $(W - S)$ ,  $(W - A)$ ,  $(W - P)$  with  $R^2 < 0.90$  as

well as  $(n - W)$  with  $R^2 = 0.945$ , and  $(W - P)$  with  $R^2 = 0.936$ .

We present now a few QSPR studies involving physical and chemical properties of alkanes and (poly)cycloalkanes in terms of the five molecular descriptors  $Q$ ,  $S$ ,  $D$ ,  $R$ , and  $P$ . For alkanes, we examined the following physical properties: NBP (in  $^{\circ}\text{C}$ ), molar heat capacity ( $C_p$ , in  $\text{J K}^{-1} \text{mol}^{-1}$ ), standard Gibbs energy of formation in gas state ( $\Delta_f G^{\circ}_{300}$  in  $\text{kJ mol}^{-1}$ ), vaporization enthalpy ( $\Delta H_{\text{vap}}^{300}$  in  $\text{kJ mol}^{-1}$ ), refractive index at  $25^{\circ}\text{C}$  ( $n_D^{25}$ ), density at  $25^{\circ}\text{C}$  ( $\rho$ , in  $\text{kg m}^{-3}$ ), and a chemical property (octane number). Numerical data for the physical properties presented in Table 1 are as in ref 28.

Among the above properties, the normal boiling point (NBP) and the molar heat capacity ( $C_p$ ) vary appreciably with  $n$  so that isomeric alkanes  $\text{C}_3$ – $\text{C}_{10}$  cluster together, and there are no overlaps between ranges of NBP or  $C_p$  values for different isomers. As a consequence, most of the variance is already explained by  $n$ , with minor roles of branching and



**Table 4.** Research Octane Numbers (RON) for Heptane and Octane Isomers

alkanes	<i>W</i>	<i>J</i>	<i>Q</i>	<i>S</i>	<i>D</i>	<i>A</i>	<i>P</i>	RON(exp)
heptane	56	2.4475	91	10.831	5.118	11.1500	7.09206	0.00
2-methylhexane	52	2.6783	101	10.045	5.056	11.4833	6.83472	42.4
3-methylhexane	50	2.8318	107	9.867	5.054	11.6167	6.78577	52.0
3,3-dimethylpentane	44	3.3604	137	8.727	4.931	12.7500	6.48502	80.8
2,4-dimethylpentane	48	2.9532	117	9.095	4.941	11.8333	6.47502	83.1
2,2-dimethylpentane	46	3.1545	125	9.009	4.959	12.0833	6.47022	92.8
3-ethylpentane	48	2.9923	117	9.080	4.924	11.7500	6.46178	65.0
2,3-dimethylpentane	46	3.1442	125	8.958	4.944	12.0000	6.44164	91.1
2,2,3-trimethylbutane	42	3.5412	153	7.898	4.767	12.500	5.98502	112.1
octane	84	2.5301	140	13.477	5.846	13.7429	8.39871	-19.0
2-methylheptane	79	2.7158	152	12.673	5.803	14.1000	8.15952	21.7
3-methylheptane	76	2.8621	160	12.473	5.804	14.2667	8.11364	26.8
4-methylheptane	75	2.9196	164	12.391	5.800	14.3167	8.08953	26.7
2,2-dimethylhexane	71	3.1118	180	11.613	5.743	14.7667	7.83266	72.5
2,5-dimethylhexane	74	2.9278	170	11.710	5.712	14.4667	7.83117	55.5
2,4-dimethylhexane	71	3.0988	178	11.571	5.724	14.6500	7.81045	65.2
2,3-dimethylhexane	70	3.1708	182	11.534	5.730	14.7333	7.80573	71.3
3-ethylhexane	72	3.0744	176	11.592	5.700	14.4833	7.79459	33.5
3-ethyl-2-methylpentane	67	3.3549	196	11.291	5.733	15.0333	7.74181	87.3
3,4-dimethylhexane	68	3.2925	194	11.301	5.708	14.8667	7.72596	76.3
2,3,4-trimethylpentane	65	3.4642	210	10.477	5.593	15.1667	7.40006	102.7
2,2,4-trimethylpentane	66	3.3889	210	10.489	5.589	15.1667	7.39756	100.0
3,3-dimethylhexane	67	3.3734	202	10.544	5.568	14.9167	7.39678	75.5
2,2,3-trimethylpentane	63	3.6233	222	10.374	5.606	15.4167	7.38084	109.6
3-ethyl-3-methylpentane	64	3.5832	220	10.382	5.583	15.2500	7.36515	80.8
2,3,3-trimethylpentane	62	3.7083	234	10.222	5.663	15.5000	7.32098	106.1

cyclic patterns. The descriptors that have a similar variation show good correlation ability for such properties.

Intermolecular forces in alkanes and (poly)cycloalkanes do not involve hydrogen bonds, and therefore Trouton's Rule about the constancy of the vaporization entropy applies:  $T_b/\Delta_{\text{vap}}H_{300} = 21.5$ , when the normal boiling point ( $T_b$ ) is expressed in the Kelvin scale ( $T_b = \text{NBP} + 273$ ).

Among the five indices for alkanes, *S*, *D*, and *P* decrease with increasing branching and cyclicity like NBPs, index *J*, many triplet indices,<sup>29</sup> and the newly defined indices *F* and *G*.<sup>30</sup> On the other hand, *Q* and *A* increase with increasing branching and cyclicity, like  $\Delta_f G^\circ_{300}$ ,  $\rho$ , the Wiener index *W*, and the Randić index  $\chi$ .

**Correlations with Normal Boiling Points (NBPs).** Since the boiling point at normal pressure is a simple and easily available physical property that can be measured with sufficient precision, this was one of the most often investigated properties in QSPR. Generally, NBPs for hydrocarbons increase substantially with size and decrease moderately with branching and only slightly with cyclicity. Figure 5 with data from Table 1 presents plots of monoparametric correlations between NBP values of all alkanes with 3–10 carbon atoms and the five TIs and, in addition, with  $W^{1/2}$ . One can see that the plots are nonlinear, especially for lower *n* values. Because NBPs increase with *n* and decrease with branching like indices *S*, *D*, and *P*, the corresponding plots are “well-arranged”; by contrast, indices *Q* and *A* increase with branching, and therefore their plots are “bristled the wrong way”.

Whereas the  $R^2$  values for correlations with *Q* and *S* are fairly low, all the remaining plots of Figure 5 have higher  $R^2$  values: 0.92 for *A*, *P*, and  $W^{1/2}$  but 0.96 for *D*. Actually, these high  $R^2$  values for the correlations with indices *D* and *A* are mainly due to the fact that the NBP values for various *n* values are not overlapping (as it does happen for correlations with the other TIs).

Biparametric correlations raise  $R^2$  values even more, as shown by eq 1 and Figure 6 in terms of indices *P* and *J* for  $C_6$ – $C_{10}$  alkanes. Cross-validation was performed by the leave-one-out procedure. Data for these correlations are presented in the last columns of Table 1.

$$\text{NBP } (^\circ\text{C}) = 16.57(\pm 1.313)J + 17.85(\pm 0.411)P - 78.42$$

$$N = 140 \quad R^2 = 0.945 \quad \text{SD} = 2.534 \quad F = 1175.8$$

$$R^2_{\text{cross-valid.}} = 0.895 \quad (1)$$

On examining the monoparametric correlations of NBPs for 37 acyclic and (poly)cyclic saturated hydrocarbons with 4–6 carbon atoms (Table 3 and Figure 7), it was found that among all five indices, index *P* gives by far the best correlation. Interestingly, although among the five TIs index *D* presents the lowest  $R^2$  value for the monoparametric correlation, its combination with *P* affords the best among the biparametric correlations, eq 2:

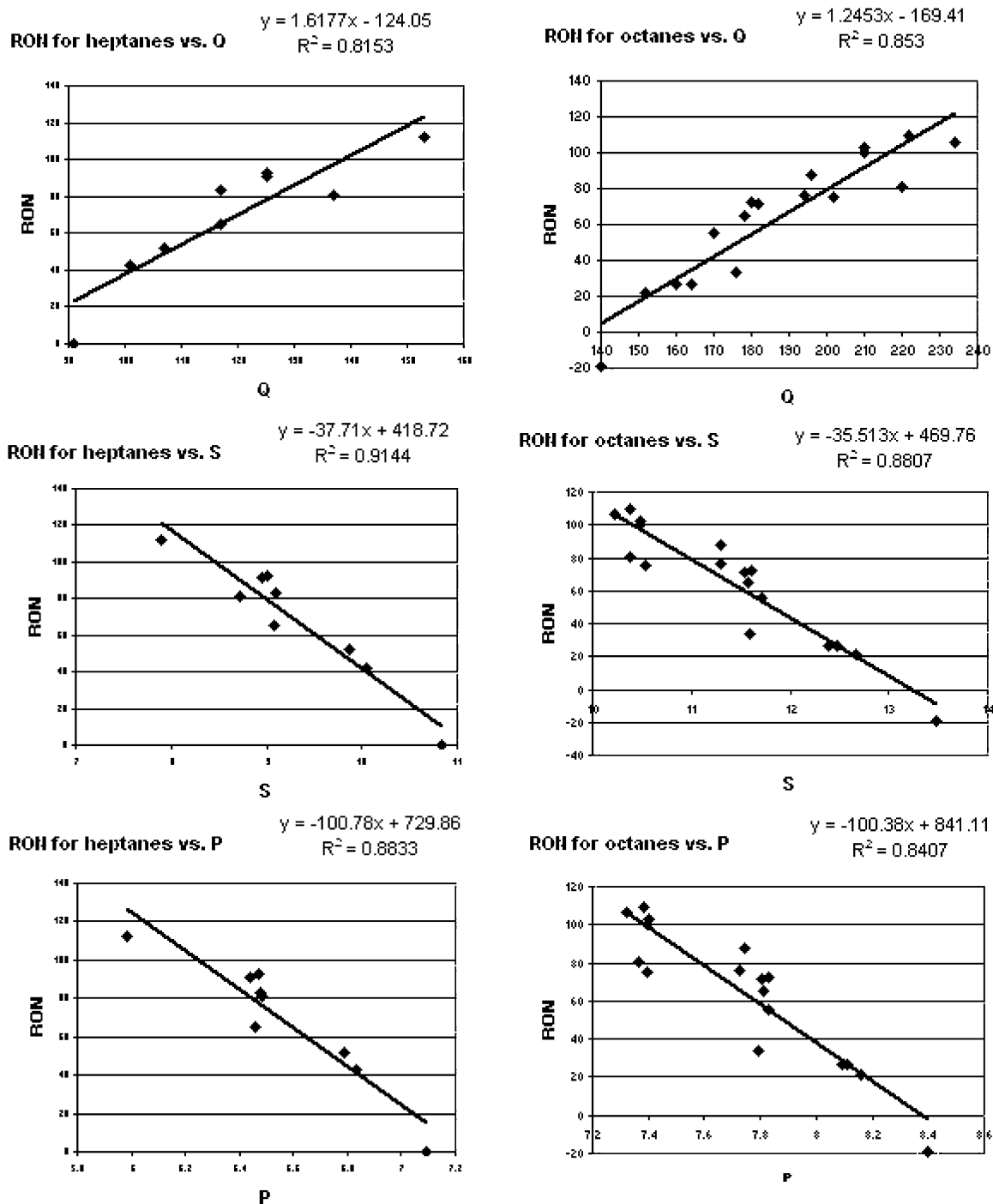
$$\text{NBP } (^\circ\text{C}) = 30.58(\pm 1.585)P - 6.477(\pm 1.738)D - 78.14$$

$$N = 37 \quad R^2 = 0.921 \quad \text{SD} = 2.717 \quad F = 198.4$$

$$R^2_{\text{cross-valid.}} = 0.867 \quad (2)$$

The last columns of Table 3 as well as Figure 8 present data and the plot for this biparametric correlation of NBPs for  $C_4$ – $C_6$  alkanes and (poly)cycloalkanes.

**Correlations with Other Physical Properties.** Five other physical properties of alkanes  $C_6$ – $C_{10}$  were investigated: specific heat ( $C_p$ , in kJ/mol), heat of vaporization ( $\Delta H_{\text{vap}}^{300}$  in kJ/mol), density ( $\rho$ , in kg/m<sup>3</sup>), index of refraction ( $n_D^{25}$ ), and Gibbs energy of formation for gaseous state,  $\Delta_f G^{300}(\text{g})$  (in kJ/mol). Numerical values of all these properties (given in Table 1) increase with *n* and decrease (not always



**Figure 11.** Plots of octane numbers RON versus TIs  $Q$ ,  $S$ , and  $P$  for heptane and octane isomers.

regularly) with branching. For four TIs, namely  $S$ ,  $D$ ,  $A$ ,  $P$  which have numerical values in the range from 4 to 24, we present in Figure 9 reverse plots of property versus these TIs: with color coding of these plots. Because index  $Q$  has for these alkanes values ranging from about 50 to 600, correlations between this TI and physical properties are presented separately in Figure 10. As shown in this figure for correlations with index  $Q$ , which increases both with  $n$  and with branching, the plot of the heat of vaporization is “bristled the wrong way”.

The specific heat is practically the same for all isomeric alkanes, and this can be seen clearly in the plots of  $C_p$  versus all TIs. It can be observed from Figures 9 and 10 that the specific heat correlations have acceptable  $R^2$  values only

because of strong correlation with  $n$ , as in the case of the correlation of  $C_p$  with  $Q$ . The correlations with the heat of vaporization  $\Delta H_{\text{vap}}^{300}$  are “bristled the wrong way”, just as in the case of correlations with NBPs. This is easily understandable (as mentioned earlier in connection with Trouton’s Rule) because both these physical properties (NBP and  $\Delta H_{\text{vap}}^{300}$ ) are due to the same weak intermolecular forces. The other three physical properties (density, Gibbs energy of formation, and index of refraction), which do not vary regularly with branching, present modest correlation coefficients.

**Correlations with Octane Numbers.** Among chemical properties of hydrocarbons, the octane numbers MON (motor octane numbers) and RON (research octane numbers)

increase with branching for the isomeric series of heptanes and octanes. Using data from ref 31, we present in Table 4 RON data for heptane and octane constitutional isomers. In Figure 11 we present monoparametric correlations between RON values and three of the new indices, namely  $Q$ ,  $S$ , and  $P$ . The other two indices ( $D$  and  $A$ ) afford lower-quality correlations. Also, combining RON values for both these classes of alkanes does not lead to a satisfactory correlation but rather to two parallel lines for the two  $n$  values.

## CONCLUSIONS

We described four new TIs based on the number of paths  $p_i$  with length increasing from  $i = 1$  (i. e. the number of edges) to the maximal value of  $p_i$ , which form the molecular path code. We had in view to include as much information as possible from this code, keeping the degeneracy to a minimum and achieving a reasonable ordering of sets of isomeric alkanes in terms of their degree of branching. For a set of isomeric (poly)cycloalkanes, branching and cyclicity cannot be varied independently, but a certain reasonable ordering can be intuitively appreciated. Among these indices,  $Q$  (an integer) and  $R$  (a real number) have appreciable degeneracy (including stochastic, apparent, and real degeneracy);  $S$  has only apparent and real degeneracy, and the lowest possible degeneracy (real degeneracy) happens when the molecular path code for two or more alkane isomers has the same sequence of numbers in the same order. This is achieved by  $A$  (which, however, does not sort higher isomeric alkanes in a reasonable order) and by  $P$ , which appears to be the best among all five indices.

Among physical properties we studied QSPR correlations with normal boiling points for all alkanes with 3–10 carbon atoms and for 37 alkanes and (poly)cycloalkanes with 4–6 carbon atoms. Correlations between chemical structure and five other physical properties or a chemical property (octane number) were also studied for alkanes.

In the future the four new TIs may be developed to be able to encode information on bond multiplicity and on the presence of heteroatoms.

**Supporting Information Available:** Constitutional isomers of alkanes with 11 and 12 carbon atoms, respectively, with their molecular path codes and selected topological indices (Tables S1 and S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- Platt, J. R. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **1947**, *15*, 419–420.
- Gordon, M.; Scantlebury, G. R. Nonrandom polycondensation: Statistical theory of the substitution effect. *Trans. Faraday Soc.* **1964**, *60*, 604–621.
- Hosoya, H. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- Gutman, I.; Rušić, B.; Trinajstić, N.; Wilcox, C. F. Graph theory and molecular orbitals. XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3409.
- Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. Molecular connectivity. Part 1. Relationship to nonspecific local anesthesia. *J. Pharm. Sci.* **1975**, *64*, 1971–1974.
- Rouvray, D. H. Graph theory in chemistry. *Roy. Inst. Chem. Rev.* **1971**, *4*, 173–195. Rouvray, D. H. *Am. Sci.* **1973**, *61*, 729–735.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Balaban, A. T. *Chemical Applications of Graph Theory*; Academic Press: London, 1976.
- Balaban, A. T. Chemical graphs. Part 35. Five new topological indices for the branching of tree-like graphs. *Theor. Chim. Acta* **1979**, *5*, 239–261.
- Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, 1984 (1st ed.); 1992 (2nd ed.).
- Topological Indices and Related Descriptors in QSAR and QSPR; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: The Netherlands, 1999.
- Bonchev, D.; Trinajstić, N. Information theory, distance matrix, and molecular branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*; Wiley – Research Studies Press: New York, 1983. Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Elsevier: Amsterdam, 1999.
- Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: New York, 2000. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.
- Randić, M. Characterization of atoms, molecules and classes of molecules based on path enumeration. *MATCH, Commun. Math. Comput. Chem.* **1979**, *7*, 3–60. Randić, M.; Wilkins, C. L. A procedure for characterization of rings of a molecule. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 36–46. Randić, M. Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 607–613.
- Randić, M.; Brisse, G. M.; Spencer, R. B.; Wilkins, C. L. Use of self-avoiding paths for characterization of graphs with multiple bonds. *Comput. Chem.* **1980**, *4*, 27–43.
- Jurs, P. C. *Computer Software Applications in Chemistry*; Wiley: New York, 1986.
- E-Dragon Software: Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual computational chemistry laboratory – design and description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453–463.
- Randić, M. On the role of mathematical chemistry in QSAR. *SAR QSAR Environ. Res.* **2006**, in press.
- Randić, M. Characterization of atoms, molecules and classes of molecules based on paths enumerations. *MATCH, Commun. Math. Comput. Chem.* **1979**, *7*, 5–64. Randić, M. On isomeric variations in decanes. *MATCH, Commun. Math. Comput. Chem.* **1982**, *13*, 271–290.
- Randić, M. Linear combinations of path numbers as molecular descriptors. *New J. Chem.* **1997**, *21*, 945–951.
- Bonchev, D. Overall connectivities/topological complexities: a new powerful tool for QSAR/QSPR. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 934–941. Bonchev, D. The overall Wiener index – a new tool for characterization of molecular topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 582–592.
- (a) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalčić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235–250. Mihalčić, Z.; Trinajstić, N. A graph-theoretical approach to structure-property relationships. *J. Chem. Educ.* **1992**, *69*, 701–712. (b) Ivanciuc, O.; Balaban, T.-S.; Balaban, A. T. Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.* **1993**, *12*, 309–318. Balaban, A. T. Local versus global (i.e. atomic versus molecular) numerical modeling of molecular graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398–402. Bonchev, D.; Balaban, A. T.; Liu, X.; Klein, D. J. Molecular cyclicity and centrality of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. *Int. J. Quantum Chem.* **1994**, *50*, 1–20. (c) Estrada, E. Generalization of topological indices. *Chem. Phys. Lett.* **2001**, *336*, 248–252.
- Balaban, A. T. A comparison between various topological indices, particularly index  $J$  and Wiener's index  $W$ . In *Topology in Chemistry: Discrete Mathematics of Molecules*; Rouvray, D. H., King, R. B., Eds.; Horwood Publishing Ltd.: Chichester, 2002; pp 89–112. Balaban, A. T.; Mills, D.; Basak, S. C. Alkane ordering as a criterion for similarity between topological indices: index  $J$  as a “sharpened Wiener index”. *MATCH Commun. Math. Comput. Chem.* **2002**, *45*, 5–26.
- Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *80*, 399–404. Balaban, A. T. Topological indices based on topological distances in molecular graphs. *Pure Appl. Chem.* **1983**, *55*, 199–206. Balaban, A. T. Chemical graphs. 48. Topological index  $J$  for heteroatom-containing molecules taking into account periodicities of element properties. *MATCH Commun. Math. Comput. Chem.* **1986**, *21*, 115–122. Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for the computation of

- molecular graph descriptors for heteroatom-containing molecules. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 395–401.
- (27) Bertz, S. H. Branching in graphs and molecules. *Discr. Appl. Math.* **1988**, 19, 65–83.
- (28) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. The complementary distance matrix, a new molecular graph descriptor. *Acta Chem. Hung., Models Chem.* **2000**, 137, 57–82. Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D.; Balaban, A. T. Evaluation in quantitative structure-property relationship models of structural descriptors derived from information-theory operators. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 631–643.
- (29) Filip, P. A.; Balaban, T. S.; Balaban, A. T. A new approach for devising local graph invariants: derived topological indices with low degeneracy and good correlation ability. *J. Math. Chem.* **1987**, 1, 61–83.
- (30) Balaban, A. T.; Mills, D.; Kodali, V.; Basak, S. C. Complexity of chemical graphs in terms of size, branching, and cyclicity. *SAR QSAR Environ. Res.* **2006**, 17, 429–450.
- (31) Balaban, A. T.; Motoc, I. Chemical Graphs. 36. Correlations between octane numbers and topological indices of alkanes. *MATCH Commun. Math. Comput. Chem.* **1979**, 5, 197–218. Balaban, A. T.; Kier, L. B.; Joshi, N. Structure-property analysis of octane numbers for hydrocarbons (alkanes, cycloalkanes, alkenes). *MATCH Commun. Math. Comput. Chem.* **1992**, 28, 13–27.

CI6005068