CROATICA CHEMICA ACTA CCACAA **69** (3) 857–871 (1996)

ISSN-0011-1643
CCA–2347

Original Scientific Paper

Molecular Topology 22.[1] Novel Connectivity Descriptors Based on Walk Degrees

Mircea V. Diudea, Ovidiu M. Minailiuc, and Gabriel Katona*

Department of Chemistry, Babes-Bolyai University of Cluj, Arany Janos 11, 3400 Cluj, Romania

Received April 6, 1995; revised August 17, 1995; accepted August 30, 1995

An algorithm for generating novel connectivity topological descriptors, denoted SP (subgraph property), is proposed and exemplified for P being the number of vertices N, walk degree $W^{(e)}$, Randić index χ, and Wiener index W. SP indices based on $W^{(e)}$ and $\chi W^{(e)}$ (Razinger's extension of χ index) are tested for correlation with some physico-chemical properties of octane isomers.

INTRODUCTION

A walk, $W^{(e)}$, in a graph G = G(V, E) ia s continuous sequence of vertices,[2] v_1, v_2, ..., v_m, its edges and vertices are allowed to be revisited. If the two terminal vertices coincide ($v_1 = v_m$), the walk is called a closed (or self returning) walk, otherwise it is an open walk. If its vertices are distinct, the walk is called a path. The number (**e**) of edges traversed is called the length of walk.

The counting of walks of length (e), starting at vertex $i \in V$(G), can be accomplished by summing the entries $[A^e]_g$ in row i of the e^{th} power of the adjacency matrix $\boldsymbol{A}$

$$W_i^{(e)} = \sum_{j \in V(G)} [A^e]_{i,j} \tag{1}$$

* Author to whom correspondence should be addressed.

$W_i^{(e)}$ is called the walk degree[3,4] of vertex i (or atomic walk count[5]). It can also be evaluated by iterative summation of the vertex degrees k_i over all neighbours, as Morgan[6] proposed for extended connectivities *ECs* (see also[5,7a]). In this respect, Diudea *et al.*[4] have proposed an algorithm which works on the connectivity matrix $\boldsymbol{C}$ (a particular case of $\boldsymbol{A}$ matrix, with $[C]_{ii} = 0$ and $[C]_{ij}$ = connectivity (conventional bond orders)). According to this algorithm, one defines a matrix $(\boldsymbol{C})\boldsymbol{W}^{(e)}$ as the sum of matrix $\boldsymbol{C}$ and a diagonal matrix $\boldsymbol{W}^{(e)}$ (of walk degress)

$$\boldsymbol{C} + \boldsymbol{W}^{(e)} = (\boldsymbol{C})\boldsymbol{W}^{(e)} \tag{2}$$

whose elements are

$$[(C)W^{(e+1)}]_{ii} = \sum_j ([C]_{ij} * [(C)W^{(e)}]_{jj}); \quad [(C)W^{(0)}]_{jj} = 1 \tag{3}$$

$$[(C)W^{(e+1)}]_{ij} = [(C)W^{(e)}]_{ij}) = [C]_{ij} \tag{4}$$

The diagonal entries $[(C)W^{(e)}]_{ii}$ are just the walk degrees $(C)W_i^{(e)}$ or simply $W_i^{(e)}$. The algorithm accounts for multiple bonds by means of the $[C]_{ij}$ entries. When the above algoritm works on a square matrix $\boldsymbol{M}$, different from matrix, it results in wlak degrees $(M)W_i^{(e)}$ weighted[8] with that »M« property (*i.e.* $M = D$, the distance; H, the reciprocal of distance, *etc.*).

Walk degrees $W_i^{(e)}$ are local (vertex) invariants (LOVIs, their sum over all vertices in the graph, $W^{(e)}$ is a graph invariant for which correlations with physico-chemical properties (*i.e.* ^{13}C NMR chemical shifts[5]) were found. They can serve as a basis for construction of other graph theoretical invariants, sometimes called topological indices. (*TI*s). We will refer here to two of the well known *TI*s: Randić index[9] χ (as its extension given by Razinger,[7b] $\chi W^{(e)}$ and Wiener index[10] W.

For reasons concerning the SP alforithm (see below), we redefine these indices as vertex invariants, whose sum gives twice the original value (and is therefore considered a global property).

Randić – Razinger index

$$\chi W_i^{(e)} = \sum_{(i,j)\in E(G)} (W_i^{(e)} * W_i^{(e)})^{-1/2} \tag{5}$$

$$2\chi W^{(e)} = 2\chi W^{(e)}(G) = \sum_i \chi W_i^{(e)} \tag{6}$$

Wiener index

$$W_i = (\mathrm{D})W_i^{(1)} \tag{7}$$

$$2W = 2W(\mathrm{G}) = \sum_i W_i \tag{8}$$

where W_i means the sum of distances from vertex i in all other vertices in G (or its »distance walk« degree[8] of length 1, $(\mathrm{D})W_i^{(1)}$).

In this paper, an algorithm for generating novel connectivity descriptors, denoted SP (subgraph property), is proposed and compared with the Randić X'/X algorithm.[11–13] Some SP descriptors are exemplified and tested, within octane isomers, for correlation with physico-chemical properties.

SP CONNECTIVITY DESCRIPTORS

SP (subgraph property) indices are built according to the following algorithm:

(i) For each edge $e \in E(\mathrm{G})$, two subgraphs, $\mathrm{S_{L,e}}$ and $\mathrm{S_{R,e}}$, which collect the vertices lying to the left and to the right of edge **e**, are defined.

(ii) Subgraph properties, $\mathrm{P(S_{L,e})}$ and $\mathrm{P(S_{R,e})}$, are calculated by summing the vertex contributions, P_i (taken as LOVIs from the global property $\mathrm{P(G)} = \Sigma_i P_i$) of all vertices **i** belonging to the given subgraph.

$$\mathrm{P(S_{L,e})} = \sum_{i \in \mathrm{S_{L,e}}} P_\mathrm{i} \tag{9}$$

$$\mathrm{P(S_{R,e})} = \sum_{i \in \mathrm{S_{R,e}}} P_\mathrm{i} \tag{10}$$

Dividing by P(G), one obtains the normalized values $\mathrm{P'(S_{L,/R,e})}$, ranging in the domain 0 to 1.

(iii) Edge contributions SP_e are evaluated as products of the normalized subgraph properties of the two corresponding subgraphs. Since a subgraph property is taken as a part of a graph property, it is obvious that only one normalized value is needed.

$$\mathrm{SP}_e = \mathrm{P'(S_{L,e})} * \mathrm{P'(S_{R,e})} = \mathrm{P'(S_{L,e})} * (1 - \mathrm{P'(S_{L,e})}) \tag{11}$$

(iv) Global SP index is calculated by summing the edge contributions for all edges in the graph

$$SP = SP(G) = \sum_{e \in E(G)} SP_e \tag{12}$$

This paper is restricted to three graphs and P being a topological property. For cycle containing graphs, a work is in progress.

TABLE I

Illustration of the SP algorithm (Graph 234M3C5; Figure 1; Property $P_i : N_i$; $W_i^{(1)}$; $\chi W_i^{(1)}$; $(D)W_i^{(1)}$ and $P(G) = \Sigma_i P_i$)

vertex	N_i	$W_i^{(1)}$	$\chi W_i^{(1)}$	$(D)W_i^{(1)}$
1	1	1	0.57735	19
2	1	3	1.48803	13
3	1	3	1.24402	11
4	1	3	1.48803	13
5	1	1	0.57735	19
6	1	1	0.57335	19
7	1	1	0.57735	17
8	1	1	0.57335	19
P(G)	8	14	7.10684	130

SN(e): 4[1/8 * 7/8] = 4[0.125 * 0.875] = 4 * 0.10938
1[1/8 * 7/8] = [0.125 * 0.875] = 0.10938
2[3/8 * 5/8] = 2[0.375 * 0.625] = 2 * 0.23438
$SN = \Sigma_e\ SN(e) = 1.01563$

$SW^{(1)}(e)$: 4[1/14 * 13/14] = 4[0.07143 * 0.92857] = 4 * 0.06633
1[1/14 * 13/14] = [0.07143 * 0.92857] = 0.06633
2[5/14 * 9/14] = 2[0.35714 * 0.64286] = 2 * 0.23358
$SW^{(1)} = \Sigma_e\ SW^{(1)}(e) = 0.79082$

$S\chi W^{(1)}(e)$: $4[0.57735 * 6.529486]/(7.10684)^2 = 4 * 0.07464$
$1[0.57735 * 6.529486]/(7.10684)^2 = 0.07464$
$2[1.64273 * 4.464106]/(7.10684)^2 = 2 * 0.23358$
$S\chi W^{(1)} = \Sigma_e S\chi W^{(1)}(e) = 0.84035$

$S(D)W^{(1)}(e)$: $4[19 * 111]/130^2 = 4[0.14615 * 0.85385] = 4 * 0.12479$
$1[17 * 113]/130^2 = 1[0.13077 * 0.86923] = 0.11369$
$2[51 * 79]/130^2 = 2[0.39231 * 0.60769] = 2 * 0.23840$
$S(D)W^{(1)} = \Sigma_e S(D)W^{(1)}(e) = 1.08964$

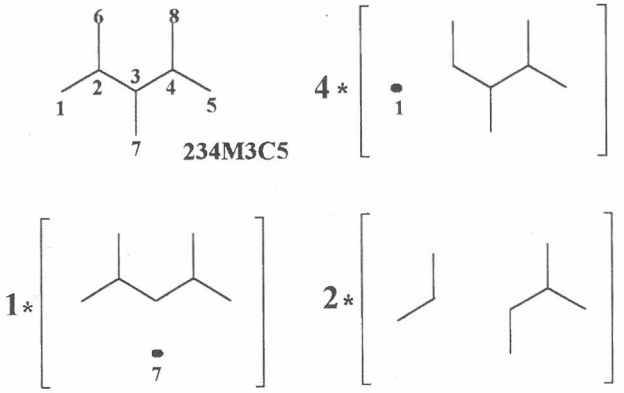

Figure 1. Subgraph fragmentation of 2,3,4-Triomethylpentane (234M3C5).

The SP algorithm is exemplified for 2,3,4-trimethylpentane: Figure illustrates subgraph fragmentation and Table I gives numerical details for four properties ($P_i = N_i$, $W_i^{(1)}$; $\chi W_i^{(e)}$, and $(D)W_i^{(1)}$).

RESULTS

SP descriptors with $P_i = W_i^{(e)}$ and $\chi W_i^{(e)}$ (*i.e.* $SW^{(e)}$ and $S\chi W^{(e)}$) and their base indices were tested for their capability to correlate with some physico-chemical properties of octane isomers. Table II collects values $W^{(e)}$ and $\chi W^{(e)}$ for octanes, *(e)* taking values from 1 to p-1.

It is interesting to note that the degeneracy within $W^{(1)}$ values gets gradually relieved until $W^{(4)}$. Clustering that appeared within $W^{(2)}$ values (marked by bars) will be considered in Discussion. Similarly, the degeneracy within $\chi W^{(1)}$ is relieved already at $\chi W^{(2)}$. Table III lists the $SW^{(e)}$ values for octane isomers. Within the $SW^{(1)}$ values, only one degeneracy appears while at higher *(e)* values it is relieved. The clustering observed within the $W^{(2)}$ data is here maintained.

One can see from Table IV that no degeneracy appears within $S\chi W^{(e)}$ values, in contrast to the basic index $\chi W^{(e)}$ (which shows two pairs of degenerate values – Table II). Base indices and the corresponding SP indices shown in Tables II to IV correlate as follows: between 0.699 ($W^{(2)}/SW^{(2)}$) and 0.898 ($W^{(3)}/SW^{(3)}$) and between 0.874 ($\chi W^{(1)}/S\chi W^{(1)}$) and 0.993 ($\chi W^{(4)}/S\chi W^{(4)}$).

Table V lists the selected physico-chemical properties of octanes. ΔHF (Heat of Formation),[14] CP (Critical Pressure),[15] ^{13}C-NMR chemical shift sum)[16] and BP (Boiling Point).[15]

Tables VI and VII offer two variable regression results.

TABLE II

Values $2W^{(e)}$ and $2\chi W^{(e)}$ for octane isomers

Graph*	$W^{(1)}$	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(5)}$	$W^{(6)}$	$W^{(7)}$	$\chi W^{(1)}$	$\chi W^{(2)}$	$\chi W^{(3)}$	$\chi W^{(4)}$	$\chi W^{(5)}$	$\chi W^{(6)}$	$\chi W^{(7)}$
C8	14	26	48	90	168	316	592	7.8284	4.2876	2.3445	1.2638	0.6792	0.3625	0.1936
2MC7	14	28	52	104	196	392	742	7.5401	3.9429	2.0839	1.0836	0.5633	0.2915	0.1506
3MC7	14	28	54	108	212	424	836	7.6161	3.9117	2.0426	1.0390	0.5300	0.2618	0.1357
4MC7	14	28	54	110	214	438	854	7.6161	3.9821	2.0477	1.0417	0.5252	0.2639	0.1327
3EC6	14	28	56	114	230	486	946	7.6921	3.9287	1.9726	0.9737	0.4827	0.2380	0.1176
25M2C6	14	30	56	118	226	468	910	7.2518	3.6038	1.8302	0.9019	0.4549	0.2252	0.1132
24M2C6	14	30	58	124	242	516	1010	7.3278	3.5988	1.7903	0.8745	0.4301	0.2102	0.1031
23M2C6	14	30	60	126	252	538	1110	7.3641	3.5673	1.7629	0.8611	0.4211	0.2034	0.0987
34M2C6	14	30	62	130	272	570	1194	7.4374	3.5068	1.7061	0.8172	0.3905	0.1866	0.0890
3E2MC5	14	30	62	132	274	582	1210	7.4374	3.5410	1.6864	0.8038	0.3827	0.1822	0.0867
22M2C6	14	32	60	138	264	606	1170	7.1213	3.5088	1.7553	0.8514	0.4188	0.2014	0.0977
33M2C6	14	32	64	146	296	676	1374	7.2426	3.4353	1.6499	0.7772	0.3642	0.1699	0.0790
234M3C5	14	32	66	144	302	656	1378	7.1068	3.1781	1.5096	0.7044	0.3302	0.1546	0.0723
3E3MC5	14	32	68	152	326	728	1562	7.3639	3.2895	1.5266	0.6969	0.3186	0.1455	0.0665
224M3C5	14	34	64	156	294	718	1354	6.8330	3.2123	1.5081	0.7079	0.3313	0.1550	0.0723
223M3C5	14	34	70	162	342	784	1666	6.9627	3.0548	1.4171	0.6453	0.2940	0.1337	0.0607
233M3C5	14	34	72	164	358	806	1770	7.0080	2.9778	1.3781	0.6161	0.2784	0.1252	0.0563
2233M4C4	14	38	80	194	434	1016	2318	6.5000	2.5535	1.1458	0.4853	0.2133	0.0919	0.0401

* M = Methyl; E = Ethyl.

TABLE III

Values $SW^{(e)}$ for octane isomers

Graph*	$SW^{(1)}$	$SW^{(2)}$	$SW^{(3)}$	$SW^{(4)}$	$SW^{(5)}$	$SW^{(6)}$	$SW^{(17)}$
C8	1.17857	1.15533	1.16632	1.11123	1.09573	1.09768	1.09033
2MC7	1.07653	1.10077	1.03402	1.06148	1.01931	1.04783	1.01219
3MC7	1.01531	1.01148	0.97154	0.97025	0.95862	0.95753	0.95300
4MC7	0.94900	0.98087	0.93519	0.94934	0.91831	0.94097	0.91396
3EC6	0.93367	0.91199	0.89094	0.89451	0.88868	0.89008	0.88819
25M2C6	0.97449	1.05444	0.96365	1.03067	0.97465	1.01977	0.98340
24M2C6	0.91327	0.96335	0.90101	0.94830	0.90190	0.94463	0.90342
23M2C6	0.89286	0.92556	0.87778	0.89298	0.87430	0.88161	0.87344
34M2C6	0.85204	0.87444	0.85770	0.85615	0.85527	0.85459	0.85472
3E2MC5	0.80612	0.82000	0.80645	0.81933	0.80641	0.81906	0.80637
22M2C6	0.91327	0.96973	0.86166	0.92297	0.84272	0.89883	0.83230
33M2C6	0.73163	0.86429	0.80566	0.83763	0.79625	0.83063	0.79992
234M3C5	0.79082	0.87598	0.82897	0.85962	0.83236	0.85801	0.83269
3E3MC5	0.77041	0.79785	0.78720	0.78640	0.78786	0.78590	0.78789
224M3C5	0.81122	0.92993	0.80273	0.92295	0.79660	0.91868	0.79295
223M3C5	0.75000	0.84170	0.77939	0.81973	0.78126	0.81450	0.78127
233M3C5	0.72959	0.18574	0.77623	0.79019	0.78361	0.78579	0.78527
2233M4C4	0.64796	0.81510	0.72906	0.78008	0.75203	0.76814	0.75911

* M = Methyl; E = Ethyl.

DISCUSSION

A single number representation of a chemical structure, in graph theoretical terms, is called a topological descriptor.[12] When a structure is »condensed« within a single number, it results in a considerable loss of information. However, such graph invariants have found large application in the explication and prediction of various molecular properties (some reviews are available[16–18]) as well as in the isomorphism and similarity studies.[3] When a topological descriptor can account for a molecular property, it can be termed a molecular index or topological index.[12] Over one hundred topological indices have been devised to date. Randić has claimed[12] some desirable attributes for TIs aimed at preventing their disliked proliferation.

TABLE IV

Values $S\chi W^{(e)}$ for octane isomers

Graph*	$S\chi W^{(1)}$	$S\chi W^{(2)}$	$S\chi W^{(3)}$	$S\chi W^{(4)}$	$S\chi W^{(5)}$	$S\chi W^{(6)}$	$S\chi W^{(17)}$
C8	1.25116	1.27773	1.29782	1.30566	1.31218	1.31367	1.31616
2MC7	1.13309	1.14934	1.16227	1.16398	1.16519	1.16454	1.16355
3MC7	1.09493	1.11339	1.13129	1.13278	1.13599	1.13540	1.13581
4MC7	1.08135	1.11226	1.12558	1.13382	1.13716	1.13927	1.14010
3EC6	1.03123	1.04885	1.05758	1.05644	1.05879	1.05790	1.05887
25M2C6	1.01173	1.01711	1.02560	1.02131	1.02545	1.02208	1.02481
24M2C6	0.97479	0.97927	0.98841	0.98687	0.98946	0.98824	0.98926
23M2C6	0.96260	0.97680	0.98800	0.98695	0.99099	0.98913	0.99082
34M2C6	0.93526	0.94134	0.94896	0.95129	0.95159	0.95219	0.95212
3E2MC5	0.88584	0.88846	0.88722	0.88829	0.88781	0.88820	0.88810
22M2C6	0.97274	0.99128	1.00269	0.99753	0.99963	0.99403	0.99387
33M2C6	0.91980	0.93764	0.94816	0.95153	0.95300	0.85371	0.95383
234M3C5	0.84035	0.83282	0.83727	0.83644	0.83690	0.83686	0.83678
3E3MC5	0.87098	0.68768	0.87340	0.87318	0.87346	0.87345	0.87357
224M3C5	0.84762	0.85031	0.85049	0.85105	0.85063	0.85062	0.85020
223M3C5	0.80972	0.81066	0.81421	0.81677	0.81621	0.81734	0.81697
233M3C5	0.79633	0.78895	0.79749	0.79669	0.79744	0.79741	0.79740
2233M4C4	0.67604	0.66119	0.66970	0.66494	0.66761	0.66603	0.66695

* M = Methyl; E = Ethyl.

Keeping Randić's suggestions in mind we propose here an algorithm which offers topological descriptors, generic term SP (subgraph property), as a third approach (along with the χ-type and X'/X-type ways) for devising connectivity indices.

We started from Wiener's original procedure[10] for calculating his W number in acyclic graphs: by summing the bond contributions W. These are obtained by multiplying the number of vertices $N_{L,e}$, on the left side, by the number of vertices $N_{R,e}$, on the right side of edge e, or also by $N - N_{L,e}$ (N being the number of vertices in the whole graph)

$$W_e = N_{L,e} * N_{R,e} = N_{L,e} * (N - N_{L,e}) \tag{13}$$

$$W = \Sigma_e We \tag{14}$$

TABLE V

Physico-Chemical properties of octanes: ΔHF (Heat of Formation[14]); CP (Critical Pressure[15]); ^{13}C (^{13}C-NMR chemical shift sum[16]) and BP (Boiling Point[15])

Graph*	ΔHF	CP	^{13}C	BP
C8	–49.90	24.54	195.6	125.8
2MC7	–51.47	24.52	208.4	117.6
3MC7	–50.79	25.33	199.2	118.8
4MC7	–50.66	25.09	198.6	117.7
3EC6	–50.36	25.74	182.5	118.9
25M2C6	–53.18	24.54	220.2	108.4
24M2C6	–52.40	25.23	209.4	109.4
23M2C6	–51.10	25.94	195.6	115.3
34M2C6	–50.87	26.57	183.5	118.7
3E2MC5	–50.44	26.65	184.6	115.6
22M2C6	–53.67	24.96	226.4	107.0
33M2C6	–52.58	26.19	204.6	112.0
234M3C5	–51.19	26.94	194.3	133.4
3K3MC5	–51.35	27.71	172.3	118.2
224M3C5	–53.54	25.34	248.6	99.3
223M3C5	–52.57	26.94	210.4	110.5
233M3C5	–51.69	27.83	191.3	114.6
2233M4C4	–53.95	28.30	223.6	106.0

* M = Methyl; E = Ethyl.

Disregarding the normalizing factor, the relatedness of eqs (11), (12) and (13), (14) is straightforward. Thus, it is obvious that the SN index is just the Wiener index, divided by squared N (Table I). When P is a property other than N, the correlation between the basic property (*i.e.* a topological index) and the derived SP descriptor is less than unity (see above).

The product operation in eq (11) induces one of the essential differences between the SP and X'/X descriptors. This results in a lower degeneracy of SP indices (Tables II to IV) by comparison with X'/X. Another difference is the mode of evaluation of P within the two subgraphs resulting from the edge cutting procedure. While the X'/X algorithm considers P as an internal property (*i.e.* the subgraph is taken as an independent graph) and hence it builds an »external« property (see the more general procedure of Mekenyan *et al.*.[19] for evaluating fragmental indices) as an edge contribution, the SP algorithm simply sums P_i (the vertex contributions to the basic graph property P(G) to evaluate the subgraph property and, next, the edge contribution.

The SP descriptors are bond additive quantities that show good correlation with some physico-chemical properties of alkanes. Since SP express the connectivity in the graph, they can be used for weighting the adjacency matrix.

TABLE VI

Two variable regression in octanes: $W^{(e)}$ and $SW^{(e)}$ *vs.* ΔHF; CP; ^{13}C and BP

Property	TI	b_i	a	R	S	F
ΔHF	$W^{(2)}$	−1.03899	−36.94523	0.95002	0.41230	69.45688
	$W^{(3)}$	0.28050				
	$SW^{(2)}$	31.21853	−52.87645	0.91897	0.52077	40.73549
	$SW^{(3)}$	−28.31920				
CP	$W^{(2)}$	−0.52857	22.52414	0.99483	0.12929	720.33584
	$W^{(3)}$	0.32243				
	$SW^{(2)}$	−14.95449	35.84332	0.91017	0.52759	36.20818
	$SW^{(3)}$	4.61777				
^{13}C	$W^{(2)}$	19.41912	6.45761	0.98076	3.87474	189.34360
	$W^{(3)}$	−6.56796				
	$SW^{(2)}$	511.83671	143.47588	0.94182	6.67204	58.89122
	$SW^{(3)}$	−474.50158				
BP	$W^{(2)}$	2.61462	78.89577	0.88622	3.05942	27.44753
	$W^{(3)}$	−0.93893				
	$SW^{(2)}$	159.61654	110.19371	0.97357	1.50845	136.25787
	$SW^{(3)}$	−150.99953				

TABLE VII

Two variable regression in octanes: $\chi W^{(e)}$ and $S\chi W^{(e)}$ *vs.* ΔHF; CP; ^{13}C and BP

Property	TI	b_i	a	R	S	F
ΔHF	$W^{(1)}$	6.79134	−90.69685	0.92252	0.50971	42.85260
	$W^{(2)}$	−3.03198				
	$SW^{(1)}$	18.25708	−57.46463	0.57562	1.07996	3.71635
	$SW^{(2)}$	−12.18966				
CP	$W^{(1)}$	5.46096	8.30571	0.96936	0.31288	116.77607
	$W^{(2)}$	−6.31053				
	$SW^{(1)}$	12.59466	31.72815	0.85022	0.67048	19.56313
	$SW^{(2)}$	−18.45508				
^{13}C	$W^{(1)}$	−152.16654	961.13316	0.98619	3.28756	265.93947
	$W^{(2)}$	100.18993				
	$SW^{(1)}$	−505.79223	252.79920	0.22257	19.35264	0.39039
	$SW^{(2)}$	444.14332				
BP	$W^{(1)}$	34.46258	−81.56376	0.90757	2.77309	35.03694
	$W^{(2)}$	−15.97579				
	$SW^{(1)}$	172.75652	79.00779	0.62474	5.15671	4.80121
	$SW^{(2)}$	−135.01293				

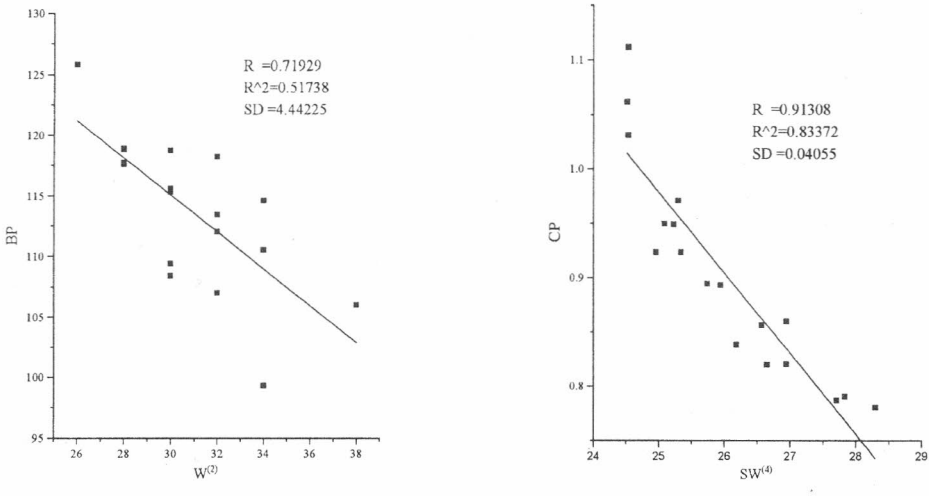

Figure 2. Plots of BP *vs.*, $W^{(2)}$ and **CP** *vs.* $SW^{(4)}$.

Edge contributions SP_e take larger values for more central edges than for the external ones, as it can be seen from Table I.

We have tested capability of SP, ($P_i = W_i^{(e)}$ and $\chi W_i^{(e)}$) and their base indices to correlate with some physico-chemical properties of octanes, in single and two variable regressions.

(i) Single variable regressions are, in general, unsatisfactory. $W^{(e)}$ and $SW^{(e)}$ show alternate values of the correlation index R, with larger values for *(e)* = even, in the case of $W^{(e)}$, and *(e)* = odd, in the case of $SW^{(e)}$. When the property is CP, the situation is reversed and when it is ^{13}C (NMR chemical shifts sum) R is larger for *(e)* = even, both for $W^{(e)}$ amd $SW^{(e)}$. Except for CP, the tested properties show values of R larger for the basic descriptors than for the corresponding SP indices, however, they are still poor. $\chi W^{(e)}$ and $S\chi W^{(e)}$ show similar trends. Figure 3 shows plots BP *vs* $W^{(2)}$ (R = 0.7193) and CP *vs.* $SW^{(4)}$ R = 0.9138), the last correlation being about 10% higher than that given by $W^{(4)}$.

(ii) Two variable regressions result in a considerable improvement of correlation. $W^{(e)}$ descriptors correlate over 0.950 with all the considered properties while $SW^{(e)}$ show lower R values (but still good – Table VI). An exception is BP, for which $SW^{(3)}$ and $SW^{(4)}$ show a correlation ($_i$0.973565) that exceeds all other correlations reported for octanes. (to our knowledge) in literature. It is about 10% higher than given by the corresponding base discriptors. As *(e)* increases, the correlation decreases. We used a regression equation of the form

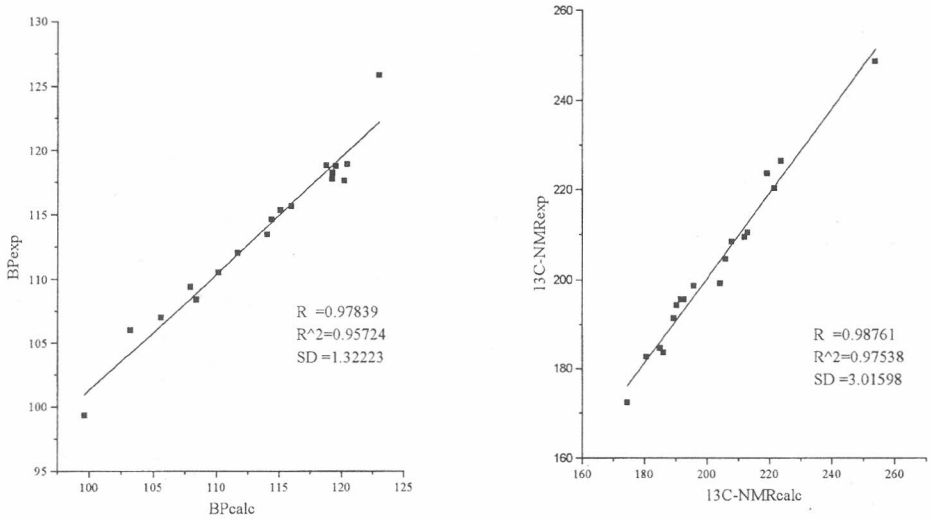

Figure 3. Plots of BP_{calc} (by Eq. (15) with $SW^{(3)}$ and $SW^{(4)}$) vs. BP_{exp} and ^{13}C-NMR_{calc} (by eq. 15 with $\chi W^{(3)}$ and $\chi W^{(4)}$) vs. ^{13}C-NMT_{exp}.

$$Y_{calc} = a + \Sigma_i \, b_i \, TI_i \qquad (15)$$

Figure 3 shows the plot BP_{calc} (by Eq. (15) with $SW^{(3)}$ and $SW^{(4)}$) against experimental BP.

$\chi W^{(e)}$ and $S\chi W^{(e)}$ show correlations (Table VII) which are, ion general, lower than those obtained for $W^{(e)}$ and $SW^{(e)}$ (except for the case of ^{13}C) and much higher for the basic $\chi W^{(e)}$ descriptors than for the $S\chi W^{(e)}$ ones. Figure 3 also shows the plot $^{13}C_{calc}$ (with $\chi W^{(2)}$ and $\chi W^{(3)}$) against experimental ^{13}C.

An interesting result is that the $W^{(2)}$ and $W^{(3)}$ descriptors show exactly the same statistics of correlation (except for the regression coefficients) as $P^{(2)}$ and $P^{(3)}$ descriptors, (path of length 2 and 3, respectively) cited by Randić in several works.[12,16,20] The author found that alkanes over C8 show interesting clustering of their properties (physico-chemical or mathematical). He drew $P^{(2)}/P^{(3)}$ grids for bidimensional embedding of alkane properties. Indeed, the herein selected properties show high regularity (and the corresponding high correlation – Tables VI and VII) within these coordinates.

Based on our earlier results[3,4] and the others[5] regarding the ability of $W^{(e)}$ to describe the molecular structures and to correlate with physico-chemical properties of alkanes, we propose here a new grid in $W^{(2)}/W^{(3)}$ coordinates (Figure 4). The clustering suggested by $W^{(2)}$ (and $W^{(3)}$) in Table II is confirmed by the set of $SW^{(e)}$ values in octanes (Table III) and supported by the high correlations obtained within these coordinates (Table VI).

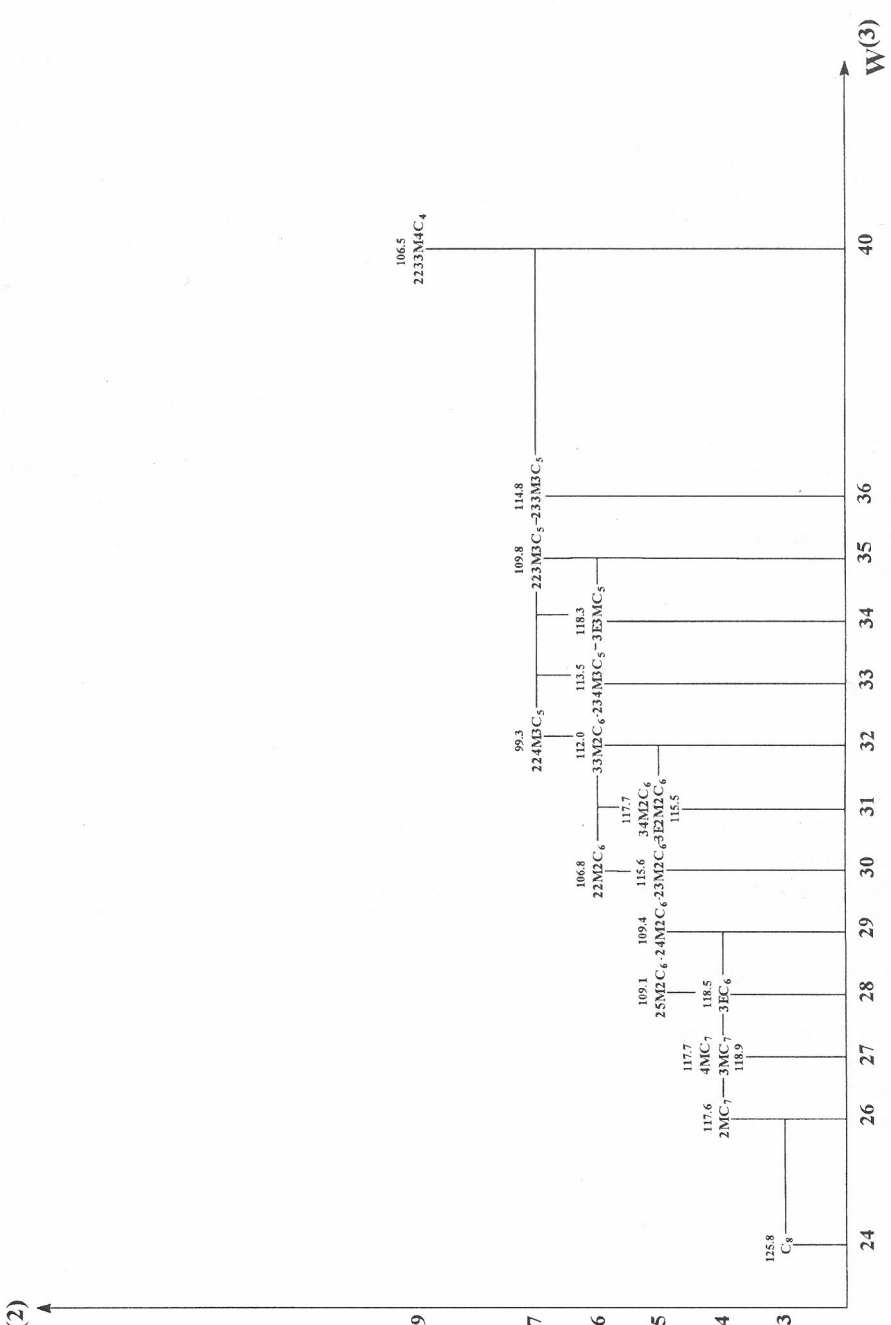

Figure 4. Ordering of octane isomers on the $W^{(2)}/W^{(3)}$ coordinates; Boiling Points.

CONCLUSIONS

SP descriptors are built by a simple algorithm that remembers the original algorithm of Wiener for calculating his W index. They are edge additive quantities, the bond contributions of which represent new graphical bond orders. The SP algorithm is the third general procedure, along with χ and X'/X algorithms, for generating connectivity indices.

Among with the SP descriptors tested herein, particular attention should be paid to $SW^{(e)}$ descriptors, especially $SW^{(3)}$ and $SW^{(4)}$, which, in the two variable regression, show a correlation of 0.9736 with the octane boiling points, the best result reported in literature to our knowledge. Preliminary results in our laboratory indicate that the $SW^{(e)}$ descriptors are far more appropriate for describing the boiling points of alkanes, than several other known topological indices (*i.e.* $P^{(e)}$, $W^{(e)}$, χ, *etc.*). SP descriptors will extend the basis for the orthogonalization procedure[20] developed by Randić for improving regression analysis in QSPR/QSAR studies. They show a lower degeneraracy in comparison with χ and X'/X indices.

From the above presented data, we can say that the SP indices obey the majority of Randić's requirements for topological indices.

Acknowledgement. – The authors are grateful to the referees for their useful comments.

REFERENCES

1. M. V. Diudea, *Wiener Index of Dendrimes* (Part 21), *MATCH* (in press).
2. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, Florida, 1983.
3. M. V. Diudea, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1064–1071.
4. M. V. Diudea, M. Topan, and A. Graovac, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1072–1078.
5. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.* **33** (1993) 683–695.
6. H. Morgan, *J. Chem. Doc.* **5** (1965) 107–113.
7. (a) M. Razinger, *Theor. Chim. Acta* **61** (1982) 581–586. (b) *Theor. Chim. Acta* **70** (1986) 365–378.
8. M. V. Diudea, *Novel Schultz Analogue Indices*, *MATCH* (in press).
9. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
10. H. Wiener, *J. Am. Chem. Soc.* **69** (1947) 17–20.
11. M. Randić, Z. Mihalić, S. Nikolić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **34** (1994) 403–409.
12. M. Randić, *J. Math. Chem.* **7** (1991) 155–168.
13. M. Randić, *Croat. Chem. Acta* **64** (1991) 43–54.
14. M. Randić, *J. Molec. Struct. (Theochem)* **232** (1991) 45–56.
15. A. T. Balaban and C. Catana, *J. Comput. Chem.* **14** (1993) 155–160.
16. M. Randić, *Concepts and Applications of Molecular Similarity*, M. A. Johnson and G. M. Maggiora (Eds.), John Wiley & Sons, Inc. 1990, p.77.

17. A. T. Balaban, I. Motoc, D. Bonchev, and O. Mekenyan, *Topological Indices for Structure – Activity Correlations, Top. Curr. Chem.* **114** (1993) 21–55.
18. D. H. Rouvray, *Chemical Applications of Topological and Graph Theory*, R. B. King (Ed.), Elsevier, Amsterdam, 1983, p. 159.
19. O. Mekenyan, D. Bonchev, and A. T. Balaban, *J. Math. Chem.* **2** (1988) 347–375.
20. M. Randić, *New J. Chem.* **15** (1991) 517–528.

SAŽETAK

Molekulska topologija 22. Novi deskriptori povezanosti zasnovani na stupnjevima šetnji

Mircea V. Diudea, Ovidiu M. Minailiuc i Gabriel Katona

Predložen je algoritam za generiranje novih topoloških deskriptora povezanosti, SP, gdje je P svojstvo podgrafa, koji je ispitan za slučaj gdje je P broj čvorova N, stupanj šetnje W^e, Randićev indeks χ i Wienerov indeks W. Korelacija SP indeksa zasnovanih na $W^{(e)}$ i $\chi W^{(e)}$ (Razingerovo proširenje indeksa χ) s fizikalno-kemijskim svojstvima testirana je na primjeru izomera oktana.