

Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes[†]

Ernesto Estrada[‡]

Departamento de Diseño de Farmacos, Centro de Bioactivos Químicos, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba

Received December 26, 1995[⊗]

A novel graph theoretical invariant based on the spectral moments of the edge adjacency matrix (**E**) is proposed. Spectral moments of the **E** matrix are used to describe seven physical properties of alkanes. All the regression models found are very significant from the statistical point of view. The spectral moments are expressed as linear combinations of the different structural fragments of the molecular graph. The use of the substructural approach for the description of seven physical properties of alkanes is also proved. The results obtained are interpreted in term of structural features of molecules.

1. INTRODUCTION

One of the main objectives of graph theoretical representation of molecules is the generation of graph invariants. These quantities, that are independent of the choice of labels for vertices (or edges) in the graph, can be used as molecular descriptors in structure–property studies.¹

Representation of molecular structures by a set of numbers (graph invariants) reduces the problem of quantitative structure–property relationships (QSPR) to a correlation between two sets of numbers via an algebraic expression. When the graph invariant is a single number such as a topological index, a considerable loss of information appears, and frequently we need to use either a convoluted algebraic expression or a combination of different single descriptors in the correlation.²

There are two possible solutions to avoid this loss of information in graph theoretical descriptors. They are (i) generalization of single descriptors to “higher” analogues or (ii) generation of graph invariants as sequences of numbers.³

Randić has made important contributions to both approaches. Among his many contributions to chemical graph theory, the initially called branching index⁴ and subsequently renamed by Kier *et al.*⁵ as the connectivity index has been generalized to a set of “higher” connectivity indices. This approach is the most successful and widespread scheme currently in use in graph theoretical QSPR studies. On the other hand, Randić has considered the sequences of path numbers,^{7,8} self-avoiding paths,^{10,11} random walks,¹² weighted path numbers,^{13,14} and basic graphs¹⁵ as approaches in composing atomic or molecular codes to be used in QSPR, QSAR, and molecular similarity studies.

Both approaches have their strong and weak points: connectivity indices are conceptually simple and easy to compute, but they have limitations when accounting for some specific branching features of molecules. In order to avoid this problem some “path-cluster” and “cluster” indices have been considered.² However, the complexity of some of the equations obtained by using these indices is forbidding.² On

the other hand, some “accidental” degeneracy has been observed in the connectivity index as a consequence of the exponent $-1/2$ used in the graph theoretical invariant.¹⁶ In a review on “third generation topological indexes” Balaban¹⁷ has shown that there is a tendency in the construction of novel topological indices with the incorporation of more electronic information about molecules into the graph theoretical invariants.

The essential difficulty of graph invariants based on sequences of numbers is related to the computational complexity of some calculations. For instance, the count of paths in polycyclic structures becomes very involved, and its proliferation in complete graphs increases exponentially with the number of vertices.¹⁸ The enumeration of the recently proposed prime path in large graphs remains also tedious and error prone, which limits its applications in such kind of graphs.¹⁵

However, chemical graph theory is continuously evolving, and novel approaches have appeared as solutions to those difficulties. Among these recent approaches we will mention the use of edge adjacency relationships^{19,20} in the generation of new topological indices related to molecular volume. Such relationships will be applied in the present paper in order to generate a series of molecular descriptors to be used in QSPR and QSAR studies.

2. SPECTRAL MOMENTS OF EDGE ADJACENCY MATRIX

Let $G = (V, E)$ be a molecular graph, with $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{e_1, e_2, \dots, e_m\}$ being the vertex- and edge-sets of G , respectively. Then the vertex-adjacency matrix **A**, commonly known as adjacency matrix, is a square and symmetric matrix, whose nondiagonal entries (i, j) are ones or zeros depending on whether vertices i and j are adjacent or not. The edge-adjacency matrix **E** of G is a square and symmetric matrix whose elements e_{ij} are 1 if and only if edge i is adjacent to edge j . Two edges are adjacent if they are incidents to a common vertex. The edge-adjacency matrix has been considered and explicitly defined in the chemical graph theory literature.^{21,22} However, this matrix has received very little attention in both chemical and mathematical literature. Recently, the present author¹⁹ has

[†] Dedicated to Professor Milan Randić for his many contributions to Chemical Graph Theory.

[‡] E-mail address: farmaco@cbqvc.sld.cu.

[⊗] Abstract published in *Advance ACS Abstracts*, June 15, 1996.

rediscovered the edge-adjacency matrix as an important source of graph theoretical invariants useful in the generation of new molecular descriptors.^{19,20} We also pointed out that the edge-adjacency matrix of a molecular graph G is identical to the vertex-adjacency matrix of the line graph of G .²³

Spectral moments of \mathbf{E} matrix are defined as follows:

$$\mu_k = \text{tr}(\mathbf{E}^k)$$

where μ_k is the k th spectral moment of \mathbf{E} and tr is the trace of the matrix. The first spectral moment μ_0 is the number of edges in the graph, because it is defined as the trace of the identity matrix of order m , where m is the number of edges. On the other hand, the second moment μ_1 is equal to zero in simple graphs, i.e., graphs having no loops.

The r th spectral moment of the edge adjacency matrix has a simple graph theoretical interpretation. It is the sum of all self-returning walks of length r in the line graph of the molecular graph, beginning and ending with the same vertex.²³

The analogous concept of spectral moments of vertex-adjacency matrix has been discussed by different authors.^{24–26} Several relations between these moments and the structures of alternant molecules has been reported in chemical literature.^{24,25} Some of these relations hold for the spectral moments of edge-adjacency matrix too, especially those obtained for simple cycles for which both matrices (\mathbf{A} and \mathbf{E}) are the same. A most detailed study of spectral moments of edge matrix for cyclic graphs will appear in a forthcoming paper.²⁷

Spectral moments of edge adjacency matrix for all alkanes (4-trees) with 3–9 carbon atoms were calculated. The values of the first 10 spectral moments for such molecules are presented in Table 1.

3. DESCRIPTION OF PHYSICAL PROPERTIES OF ALKANES

In order to test the applicability of the spectral moments of edge adjacency matrix in the search of structure–property correlations, we select seven representative physical properties of alkanes: boiling point (bp), molar volume at 20 °C (MV), molar refraction at 20 °C (MR), heat of vaporization at 25 °C (HV), critical temperature (TC), critical pressure (PC), and surface tension at 20 °C (ST). These properties were examined by Needham *et al.*²⁸ with molecular modeling techniques by using Wiener indices, connectivity indices and *ad hoc* descriptors. Most of these properties were well correlated ($r > 0.99$) with connectivity indices and *ad hoc* descriptors.

Using multiple regression by stepwise method with forward search, the first 10 spectral moments were used as independent variables and regressed against the experimental values of the physical properties. The quality of a model was determined by examining the correlation coefficient, standard deviations of regression, Fisher ratios, and the number of variables in the equation. The best linear models found are presented in Table 2.

As can be appreciated from the statistical parameters of the regression equations in Table 2, most of the physical properties are well described by the spectral moments of edge adjacency matrix. However, these equations cannot be considered as optimal. For instance, the standard deviation

in the estimation of boiling points, $s = 4.32$, is significantly higher than those obtained from the use of other topological and *ad hoc* descriptors.²⁸ The use of some topological indices based on matrix-vector multiplication by Balaban and co-workers,²⁹ has also produced low standard deviations for the description of boiling points of alkanes. These results are encouraged by the fact that as vectors one can select graph theoretical, quantum chemical, or empirical quantities. The high standard deviation obtained for some properties in the present approach can be explained considering that equations in Table 2 are completely linear, and no term accounting for nonlinear dependence among properties and spectral moments was accomplished. The generation of optimal equations for the description of physical properties is not the main objective of the present work. However, in Table 3 we illustrate the improvements produced by the introduction of the square root of μ_0 in the models to describe boiling points, critical temperatures, and critical pressures. Improvements are significant, especially for boiling points where the standard deviation was reduced to a half of the precedent value.

Almost all equations in Tables 2 and 3 are now statistically similar to those obtained by Needham *et al.* by using connectivity indices and *ad hoc* descriptors. This is well appreciated in Table 4 in which the statistical parameters for the best regression equations obtained with the moments of edge-adjacency matrix are compared to those obtained by Needham *et al.*²⁸ In this table we can observe that the standard deviations for the models obtained with moments of the \mathbf{E} matrix to describe molar refractions, heat of vaporization, and critical pressures of alkanes are the same as those obtained by using molecular connectivity indices. The equations describing molar volume and superficial tensions with the present approach have only slightly higher standard deviations than the best equations found by Needham *et al.* Only the models found by us to describe boiling points and critical temperatures have significant differences with the precedent models obtained with connectivity or *ad hoc* descriptors. However, the quality of quantitative structure–property models should not be measured only by statistical criteria. The easy interpretation of equations found in term of structural fragments of the molecules should be included as a very important criterion to select the best model in a QSPR study. Models found in the present work combine very well these two aspects, i.e., they have significant statistical parameters and easy structural interpretation, and as a consequence they can be considered as good QSPR models.

4. STRUCTURAL SIGNIFICANCE OF THE PRESENT APPROACH

Topological descriptors have been rationalized by Randić,³ who proposed a series of desirable attributes that these indices need to have. The first desirable requisite for a novel topological index is its direct structural interpretation. The discovery of the connection existing among topological descriptors and structural concepts will help one to interpret convoluted and complex physical or pharmacological properties in terms of the structure.

The spectral moments of the \mathbf{E} matrix can be expressed as linear combinations of the number of times that the different structural fragments appear as subgraphs in the

Table 1. Spectral Moments of Edge Adjacency Matrix for C₃–C₉ Alkanes

no.	alkane	m_0	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
1	3	2	2	0	2	0	2	0	2	0	2
2	4	3	4	0	8	0	16	0	32	0	64
3	2M3	3	6	6	18	30	66	126	258	510	1026
4	5	4	6	0	14	0	36	0	94	0	246
5	2M4	4	8	6	28	40	116	210	516	1032	2368
6	22MM3	4	12	24	84	240	732	2184	6564	19680	59050
7	6	5	8	0	20	0	56	0	164	0	488
8	2M5	5	10	6	34	40	142	224	642	1176	3010
9	3M5	5	10	6	38	50	172	308	846	1734	4330
10	22MM4	5	14	24	98	270	878	2646	8258	25350	78430
11	23MM4	5	12	12	52	100	300	700	1892	4692	12250
12	7	6	10	0	26	0	76	0	234	0	740
13	2M6	6	12	6	40	40	162	224	720	1194	3352
14	3M6	6	12	6	44	50	198	322	980	1896	5082
15	3E5	6	12	6	48	60	234	420	1248	2634	6972
16	22MM5	6	16	24	104	270	910	2688	8528	26050	81260
17	23MM5	6	14	12	62	110	362	826	2358	5880	15950
18	24MM5	6	14	12	54	80	254	476	1318	2784	7214
19	33MM5	6	16	24	112	300	1030	3136	10110	31700	100800
20	223MMM4	6	18	30	126	360	1230	3906	12870	41740	136500
21	8	7	12	0	32	0	96	0	304	0	992
22	2M7	7	14	6	46	40	182	224	790	1194	3614
23	3M7	7	14	6	50	50	218	322	1058	1914	5434
24	4M7	7	14	6	50	50	224	336	1114	2058	5844
25	3E6	7	14	6	54	60	260	434	1390	2814	7834
26	22MM6	7	18	24	110	270	930	2688	8614	26110	81800
27	23MM6	7	16	12	68	110	388	840	2500	6078	16900
28	24MM6	7	16	12	64	90	310	574	1664	3540	9476
29	25MM6	7	16	12	60	80	268	448	1284	2424	6396
30	33MM6	7	18	24	118	300	1062	3178	10390	32440	103800
31	34MM6	7	16	12	72	120	424	952	2832	7104	19860
32	23ME5	7	16	12	72	120	430	966	2896	7284	20440
33	33ME5	7	18	24	126	330	1188	3654	12130	38760	126200
34	223MMM5	7	20	30	136	370	1298	4074	13580	44090	145300
35	224MMM5	7	20	30	124	310	1028	2982	9444	28760	90100
36	233MMM5	7	20	30	140	390	1388	4438	14980	49350	164600
37	234MMM5	7	18	18	86	170	558	1386	4102	11030	31600
38	2233(M)4	7	24	48	204	660	2436	8652	31520	114200	416000
39	9	8	14	0	38	0	116	0	374	0	1244
40	2M8	8	16	6	52	40	202	224	860	1194	3866
41	3M8	8	16	6	56	50	238	322	1128	1914	5696
42	4M8	8	16	6	56	50	244	366	1192	2076	6196
43	3E7	8	16	6	60	60	280	434	1468	2832	8196
44	4E7	8	16	6	60	60	286	448	1532	2994	8706
45	22MM7	8	20	24	116	270	950	2688	8684	26110	82070
46	23MM7	8	18	12	74	110	408	840	1578	6096	17260
47	24MM7	8	18	12	70	90	336	588	1798	3702	10250
48	25MM7	8	18	12	70	90	324	546	1622	3144	8488
49	26MM7	8	18	12	66	80	288	448	1346	2388	6498
50	33MM7	8	20	24	124	300	1082	3178	10480	32500	104400
51	34MM7	8	18	12	78	120	450	966	2774	7302	20810
52	35MM7	8	18	12	74	100	366	672	2010	4296	11750
53	44MM7	8	20	24	124	300	1094	3220	10670	33180	106900
54	23ME6	8	18	12	78	120	456	980	3046	7500	21490
55	24ME6	8	18	12	74	100	372	686	2082	4494	12420
56	33ME6	8	20	24	132	330	1220	3696	12410	39530	129500
57	34ME6	8	18	12	82	130	492	1092	3378	8544	24540
58	223MMM6	8	22	30	142	370	1324	4088	13730	44350	146600
59	224MMM6	8	22	30	134	320	1084	3080	9798	29570	92690
60	225MMM6	8	22	30	130	310	1036	2912	9186	27390	85160
61	233MMM6	8	22	30	146	390	1420	4480	15270	50140	168000
62	234MMM6	8	20	18	96	180	620	1512	4584	12310	35820
63	235MMM6	8	20	18	88	150	500	1092	3192	7776	21600
64	244MMM6	8	22	30	138	340	1180	3472	11310	35200	113000
65	334MMM6	8	22	30	150	400	1456	4606	15700	51760	173800
66	33EE5	8	20	24	140	360	1352	4200	14300	46540	154800
67	223MME5	8	22	30	146	380	1372	4256	14370	46690	155100
68	233MME5	8	22	30	154	420	1552	4998	17250	57680	196000
69	234MEM5	8	20	18	96	180	632	1554	4776	12960	38060
70	2233(M)5	8	26	48	218	690	2600	9240	34030	124100	456400
71	2234(M)5	8	24	36	160	430	1500	4690	15700	51260	170600
72	2244(M)5	8	26	48	194	540	1808	5544	17950	56860	183500
73	2334(M)5	8	24	36	168	480	1752	5796	20230	69120	239100

Table 2. Multiple Regression Equations for the Physical Properties Using the Spectral Moments of Edge Matrix

bp (°C) = -76.7187 + 23.992 m_0 + 2.5064 m_2 - 2.9671 m_3 + 0.1492 m_5 $r = 0.9949$ $s = 4.21$ $F = 1650$
MV (cm ³) = 38.988 + 27.016 m_0 - 4.404 m_2 + 2.579 m_3 - 0.381 m_4 - 0.017 m_5 $r = 0.9993$ $s = 0.6$ $F = 9659$
MR (cm ³) = 5.703 + 5.506 m_0 - 0.329 m_2 + 0.193 m_3 - 0.033 m_4 $r = 0.9999$ $s = 0.05$ $F = 168\,569$
HV (kJ/mol) = 74.707 + 4.821 m_0 - 0.409 m_3 + 0.021 m_5 $r = 0.9988$ $s = 0.2$ $F = 9532$
TC (°C) = 77.200 + 12.145 m_0 + 12.152 m_2 - 10.598 m_3 + 1.186 m_5 - 0.182 m_6 $r = 0.9906$ $s = 7.0$ $F = 703$
PC (atm) = 48.376 - 6.605 m_0 + 1.752 m_2 - 1.270 m_3 + 0.120 m_4 + 0.101 m_5 - 0.019 m_6 $r = 0.9756$ $s = 0.8$ $F = 217$
ST (dyn/cm) = 13.054 - 0.779 m_0 + 1.058 m_2 - 0.853 m_3 + 0.049 m_4 + 0.052 m_5 - 0.006 m_6 $r = 0.9869$ $s = 0.3$ $F = 380$

Table 3. Improved Correlations for the Description of Three Physical Properties by Using a Nonlinear Term

bp (°C) = -215.72 + 120.09 (m_0) ^{0.5} + 1.61 m_2 - 2.098 m_3 + 0.029 m_5 $r = 0.9984$ $s = 2.48$ $F = 5194$
TC (°C) = -55.41 + 105.66 (m_0) ^{0.5} + 7.301 m_2 - 8.970 m_3 + 1.125 m_5 - 0.175 m_6 $r = 0.9944$ $s = 5.4$ $F = 1179$
PC (atm) = 73.509 - 24.250 (m_0) ^{0.5} + 1.575 m_2 - 1.291 m_3 + 0.006 m_4 + 0.179 m_5 - 0.031 m_6 $r = 0.9854$ $s = 0.6$ $F = 368$

Table 4. Statistical Parameters for the Models Describing Physical Properties of Alkanes by Using Connectivity Indices, ad Hoc Descriptors and Spectral Moments of Edge-Adjacency Matrix

property	connectivity indices			ad hoc descriptors			moments of E matrix		
	no. var.	r	s	no. var.	r	s	no. var.	r	s
bp	5	0.9995	1.86	5	0.9989	2.0	4	0.9984	2.48
MV	5	0.9995	0.5	5	0.9995	0.4	5	0.9993	0.6
MR	5	0.9999	0.05	5	0.9999	0.05	4	0.9999	0.05
HV	5	0.9989	0.2	5	0.9969	0.4	3	0.9988	0.2
TC	5	0.9975	4.1	5	0.9970	4.8	5	0.9944	5.4
PC	5	0.9904	0.6	5	0.9889	0.7	5	0.9854	0.6
ST	5	0.9929	0.2	5	0.9945	0.2	6	0.9869	0.3

molecular graphs. These numbers are commonly known as *embedding frequencies*^{30,31} and have been used to describe global properties of organic molecules.^{32,33}

In Table 5 we illustrate the equations relating the first eight spectral moments with the number of different fragments in the graph (embedding frequencies). The symbol $|F_i|$ represents the number of subgraphs of kind i in the molecular graph. The pictorial representations of the subgraphs of different kind are depicted in Figure 1.

In a paper of Barysz *et al.*,³⁴ the coefficients of the characteristic polynomial of vertex adjacency **A** matrix were calculated by using a similar substructural approach. However, the approach of Barysz *et al.* was not applied for the interpretation of some structural invariants used in QSPR or QSAR studies. The present approach can be used to find a connection among physical properties of alkanes and structural fragments of the molecules by using a simple method to calculate topological invariants. In Table 6 we depict the correlation equations for description of the seven studied physical properties of alkanes obtained by substituting the spectral moments of **E** matrix by its expression in terms of embedding frequencies. The statistical parameters of these equations are identical to those obtained for equations illustrated in Table 2.

Equations in Table 6 permit an easy interpretation of the physical properties in terms of the molecular structure. In order to show the influence of the individual fragments on the properties, we calculate the mean effect that each subgraph produces on the physical properties. The mean

effects, the average contribution of each fragment on the calculated physical property, are illustrated in Table 7.

The estimates of the relative contributions of molecular mass, branching, and steric factors to the physical properties observed by Needham *et al.*²⁸ are similar in global terms to the mean effects of structural fragments illustrated in Table 7. On the other hand, we can appreciate the contribution that each fragment produces on the different properties. For example, the fragment F_1 has a positive contribution to all properties with the exception of critical pressures and surface tensions; F_2 contributes positively to all properties except to those related to molecular volume (molar volume and molar refraction), while F_3 contributes positively only to MV and MR.

The present approach to quantitative structure–property relationship studies is, in some aspects, similar to the Smolenskii's additivity function.³⁵ Smolenskii's procedure is based on the decomposition of the molecular graph into different fragments. The fragments of a given type then become a variable in a multiple regression model. This approach, like the embedding frequencies, is a powerful tool in QSPR studies. The main disadvantage of the substructural approaches mentioned above is connected to the algorithmic complexity of the computation. These methods need one to generate all subgraphs in the molecular graph and then to check for isomorphism among these subgraphs. It is well-known that the complexity of such algorithms is nonpolynomial³⁶ and, of course, the number of operations in the algorithm increases very fast, e.g., exponentially, with the number of vertices in the graph. This is a real limitation of such kinds of algorithms when they are applied to molecules of some structural complexity, such as polycyclic compounds.

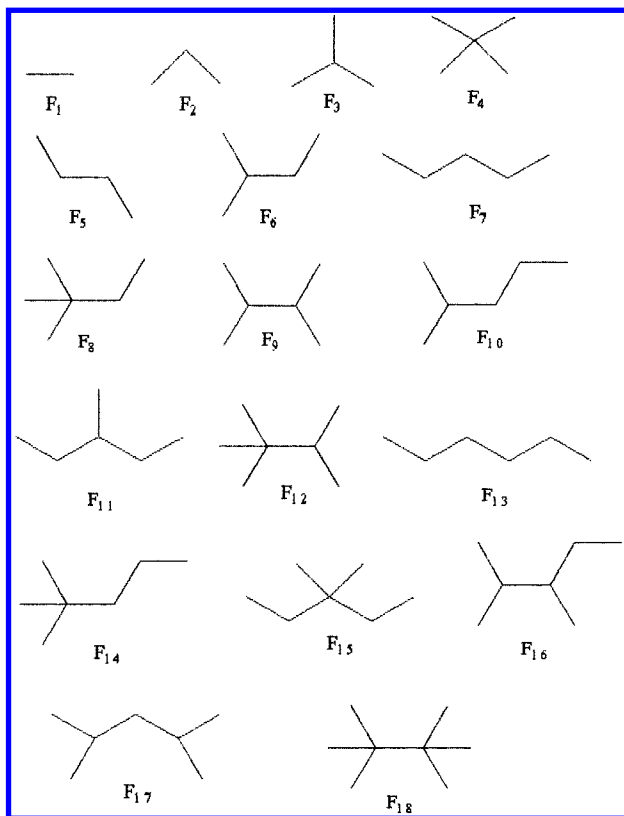
The method based on spectral moments of edge adjacency matrix also has some limitations similar to the Smolenskii-type of approaches.³⁵ The main drawback of our present approach is related to its performance in QSPR and QSAR studies when it is compared to those approaches that directly use the substructural fragments into correlations. The spectral moments are linear combinations of a series of structural fragments in the molecular graph, but not all of these fragments have a direct influence on the physical

Table 5. The First Eight Spectral Moments of **E** Matrix as Linear Combinations of the Number of Fragments in the Graph

$$\begin{aligned}
\mu_0 &= |F_1| \quad \mu_2 = 2 \times |F_2| \quad \mu_3 = 6 \times |F_3| \\
\mu_4 &= 2 \times |F_2| + 12 \times |F_3| + 24 \times |F_4| + 4 \times |F_5| \\
\mu_5 &= 30 \times |F_3| + 120 \times |F_4| + 10 \times |F_6| \\
\mu_6 &= 2 \times |F_2| + 60 \times |F_3| + 480 \times |F_4| + 12 \times |F_5| + 24 \times |F_6| + 6 \times |F_7| + 36 \times |F_8| + 24 \times |F_9| \\
\mu_7 &= 126 \times |F_3| + 1680 \times |F_4| + 84 \times |F_6| + 210 \times |F_8| + 112 \times |F_9| + 14 \times |F_{10}| + 14 \times |F_{11}| + 84 \times |F_{12}| \\
\mu_8 &= 2 \times |F_2| + 252 \times |F_3| + 5544 \times |F_4| + 28 \times |F_5| + 200 \times |F_6| + 32 \times |F_7| + 1008 \times |F_8| + 464 \times |F_9| + 32 \times |F_{10}| + 40 \times |F_{11}| + \\
&\quad 672 \times |F_{12}| + 8 \times |F_{13}| + 48 \times |F_{14}| + 46 \times |F_{15}| + 112 \times |F_{16}| + 536 \times |F_{17}| + 284 \times |F_{18}|
\end{aligned}$$

Table 6. Multiple Linear Regression Equations for the Physical Properties as Functions of Molecular Fragments

$$\begin{aligned}
\text{bp } (^{\circ}\text{C}) &= -76.719 + 23.992|F_1| + 5.01|F_2| - 13.332|F_3| + 17.880|F_4| + 1.492|F_6| \\
\text{mv } (\text{cm}^3) &= 38.988 + 27.016|F_1| - 9.570|F_2| + 10.392|F_3| - 11.184|F_4| - 1.524|F_5| - 0.17|F_6| \\
\text{MR } (\text{cm}^3) &= 5.703 + 5.506|F_1| - 0.724|F_2| + 1.050|F_3| - 0.797|F_4| - 0.120|F_5| \\
\text{HV } (\text{kJ/mol}) &= 74.707 + 4.821|F_1| - 1.813|F_3| + 2.559|F_4| + 0.213|F_6| \\
\text{TC } (^{\circ}\text{C}) &= 77.200 + 12.145|F_1| + 23.938|F_2| - 38.947|F_3| + 54.720|F_4| - 2.190|F_5| + 7.484|F_6| - 1.095|F_7| - 6.570|F_8| - 4.380|F_9| \\
\text{PC } (\text{atm}) &= 48.376 - 6.606|F_1| + 3.706|F_2| - 4.316|F_3| + 5.710|F_4| + 0.248|F_5| + 0.544|F_6| - 0.116|F_7| - 0.694|F_8| - 0.462|F_9| \\
\text{ST } (\text{dyn/cm}) &= 13.054 - 0.779|F_1| + 2.202|F_2| - 3.330|F_3| + 4.536|F_4| + 0.124|F_5| + 0.376|F_6| - 0.036|F_7| - 0.216|F_8| - 0.144|F_9|
\end{aligned}$$

**Figure 1.** First 18 structural fragments contained in molecular graphs of alkanes.

property. For instance, some fragments contained in one spectral moment may have a positive contribution to the physical property, and others may have a negative or null influence on this property. As a consequence, in such cases this spectral moment does not describe linearly the studied property and the use of nonlinear fits are necessary. However, the spectral moments of edge adjacency matrix are very easy to calculate, and they can be combined in appropriate ways in order to produce very good correlations without loss of structural information.

The structural interpretation of spectral moments of the **E** matrix appears to be the most important feature of the present approach. This is the main difference among the models found in the present paper and those obtained by Needham *et al.*²⁸ According to Mihalic and Trinajstić,² some of the QSPR models obtained with the use of higher order connectivity indices and its combinations have a forbidding complexity.

Table 7. Mean Effect of the First Six Structural Fragments on the Physical Properties of Alkanes

property	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆
bp	165.98	41.59	-37.62	7.83		7.93
MV	186.87	-79.35	28.58	-4.66	-10.33	-0.87
MR	38.08	-6.00	2.89	-0.33	-0.81	
HV	33.35		-4.99	1.07		1.09
TC	84.02	198.71	-109.91	23.97	-14.90	39.78
PC	-45.70	30.76	-12.18	2.50	1.69	2.89
ST	-5.39	18.26	-9.16	1.89	0.84	1.93

5. CONCLUSIONS

The spectral moments of the edge adjacency matrix are a novel set of topological (graph theoretical) descriptors. These indices have a series of important features that make them useful molecular descriptors to be employed in QSPR and QSAR studies. The decomposition of spectral moments of the **E** matrix into substructural fragments permits the structural interpretation of the correlation found with them to describe the physical or biological properties of organic compounds. These indices can be used as a set of descriptors in QSPR and QSAR studies in the same way as they were used in the present work. However, spectral moments of the edge adjacency matrix can also be used in combination with pattern recognition techniques and in studies of similarity/dissimilarity features of molecules.

The correlations found by using spectral moments of the **E** matrix for the description of seven representative physical properties of alkanes can be considered as statistically significant. In general, the average contributions of each descriptor to the studied physical properties are similar to the contribution of molecular mass, branching, and steric features of molecules to such properties reported by Needham *et al.*²⁸

The approach described in this paper appears to be a prominent method to find quantitative models for description of physical, thermodynamic, or pharmacological properties with a direct structural interpretation. The applications of the present method to QSAR studies of several classes of organic compounds are now in progress and will be the subject of a future publication.

REFERENCES AND NOTES

- (1) Balaban, A. T. Application of Graph Theory in Chemistry. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 334-343.
- (2) Mihalic, Z.; Trinajstić, N. A Graph Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, 69, 701-712.

- (3) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, 7, 155–168.
- (4) Randić, M. On the Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
- (5) Kier, L. B.; Hall, L. H.; Murray, W. J.; Randić, M. Molecular Connectivity I: Relationship to Nonspecific Local Anesthesia. *J. Pharm. Sci.* **1975**, 64, 1971–1975.
- (6) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
- (7) Randić, M. Characterization of Atoms, Molecules and Classes of Molecules Based on Path Enumeration. *MATCH* **1979**, 7, 5–64.
- (8) Randić, M.; Wilkins, C. L. Graph Theoretical Study of Structural Similarity in Benzomorphans. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1979**, 6, 55–71.
- (9) Wilkins, C. L.; Randić, M. A Graph Theoretical Approach to Structure–Property and Structure–Activity Correlations. *Theor. Chim. Acta* **1980**, 58, 45–68.
- (10) Randić, M.; Brissey, G. M.; Spencer, R. B.; Wilkins, C. L. Search for Self-Avoiding Path for Molecular Graphs. *Comput. Chem.* **1979**, 3, 5–13.
- (11) Randić, M.; Brissey, G. M.; Spencer, R. B.; Wilkins, C. L. Use of Self-Avoiding Path for Characterization of Molecular Graphs with Multiple Bonds. *Comput. Chem.* **1980**, 4, 27–32.
- (12) Randić, M. Random Walks and Their Diagnostic Value for Characterization of Atomic Environment. *J. Comput. Chem.* **1980**, 1, 386–399.
- (13) Randić, M. Nonempirical Approach to Structure–Activity Studies. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1984**, 11, 137–153.
- (14) Randić, M. Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure–Activity Relationships. *Chemom. Intell. Lab. Syst.* **1991**, 10, 213–227.
- (15) Randić, M. Representation of Molecular Graphs by Basic Graphs. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 57–69.
- (16) Estrada, E. Graph Theoretical Invariant of Randić Revisited. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 1022–1025.
- (17) Balaban, A. T. Using Real Number as Vertex Invariants for Third-Generation Topological Indexes. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 23–28.
- (18) Garey, M. R.; Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*; W. H. Freeman: San Francisco, 1979.
- (19) Estrada, E. Edge Adjacency Relationships and A Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 31–33.
- (20) Estrada, E. Edge Adjacency Relationships in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 701–707.
- (21) Rouvray, D. H. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; pp 180–181.
- (22) Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1983; 2nd ed.; 1992; pp 32–33.
- (23) Gutman, I.; Estrada, E. Topological Indices Based on Line Graph of the Molecular Graph. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 000–000.
- (24) Jiang, Y.; Tang, A.; Hoffmann, R. Evaluation of Moments and their Application in Hückel Molecular Orbital Theory. *Theoret. Chim. Acta* **1984**, 66, 183–192.
- (25) Cioslowski, J. Decomposition of the Total pi-electron Energy of Polycyclic Hydrocarbons into Benzene Ring Increments. *Chem. Phys. Lett.* **1985**, 122, 234–236.
- (26) Cioslowski, J. The Generalized Padé Approximants and the Chemical Graph Theory. *Int. J. Quantum Chem.* **1988**, 34, 217–224.
- (27) Estrada, E. Spectral Moments of Edge Adjacency Matrix in Molecular Graphs. 2. Graphs Containing Cycles. Manuscript in preparation.
- (28) Needham, D. E.; Wei, I.-C.; Seybold, P. G. Molecular Modeling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, 110, 4186–4194.
- (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) Poshusta, R. D.; McHughes, M. C. Embedding Frequencies of Trees. *J. Math. Chem.* **1989**, 3, 193–215.
- (31) Kvasnicka, V.; Pospíchal, J. Simple Construction of Embedding Frequencies of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 121–128.
- (32) McHughes, M. C.; Poshusta, R. D. Graph-Theoretic Cluster Expansions. Thermodynamic Properties of Alkanes. *J. Math. Chem.* **1990**, 4, 227–249.
- (33) Schmalz, T. G.; Klein, D. J.; Sandleback, B. L. Chemical Graph-Theoretical Cluster Expansion and Diamagnetic Susceptibility. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 54–57.
- (34) Barysz, M.; Knop, J. V.; Pejakovic, S.; Trinajstić, N. Characterization of Branching. *Polish J. Chem.* **1985**, 59, 405–432.
- (35) Smolenskii, E. A. *Zhur. Fiz. Khim.* **1964**, 38, 1288. Described in: Sabilij, A.; Trinajstić, N. Quantitative Structure–Activity Relationships: The Role of Topological Indices. *Acta Pharm. Jugosl.* **1981**, 31, 189–214.
- (36) Liu, X.; Klein, D. J. The Graph Isomorphism Problem. *J. Comput. Chem.* **1991**, 12, 1242–1251.

CI950187R