

Charge Indexes. New Topological Descriptors

J. Gálvez,* R. García, M. T. Salabert, and R. Soler

Unit Research into Molecular Connectivity and Drug Design, Department of Physical Chemistry,
Faculty of Pharmacy, University of Valencia, Valencia, Spain

Received February 19, 1993*

New topological descriptors, namely "charge indexes", are presented in this paper. Their ability for the description of the molecular charge distribution is established by correlating them with the dipole moment of a heterogeneous set of hydrocarbons, as well as with the boiling temperature of alkanes and alcohols and the vaporization enthalpy of alkanes. Moreover, it is clearly demonstrated that this ability is higher than that shown by the χ connectivity and Wiener indexes.

1. INTRODUCTION

A close relationship often exists between the molecular structures of organic compounds and many of their physical, chemical, or biological properties.¹⁻⁵ A great deal of these relationships have been investigated using the topological descriptors of molecular structures. These descriptors are used to characterize the constitution and configuration of a molecule by a single number. Particularly interesting are those indexes derived from the distance matrix in both topological and topographic forms. These forms have recently been studied comparatively by Mihalic et al.⁶

In this paper, we present new topological indexes, namely charge indexes, G_k and J_k , which, somehow evaluate the charge transfers between pairs of atoms, and therefore the global charge transfers in the molecule. The properties treated are the dipole moment of a mixed set of hydrocarbons, the boiling points of alkanes and alcohols, and the vaporization enthalpy of alkanes.

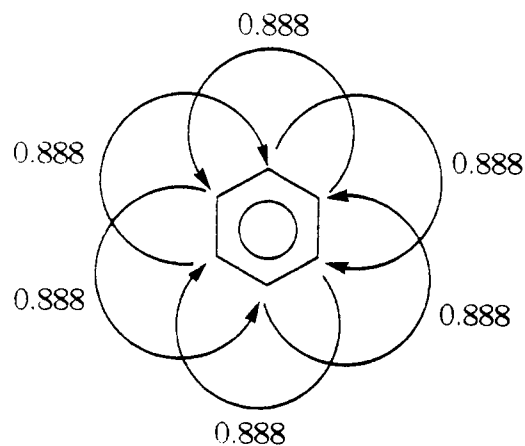
The structure of this paper will be as follows: An outline and definition of our procedures will be described in section 2. The correlation equations for the properties and molecules selected, as well as the statistical considerations on the obtained results will be developed in section 3. Finally, a discussion and the conclusions will be presented in the last section.

2. METHODOLOGY

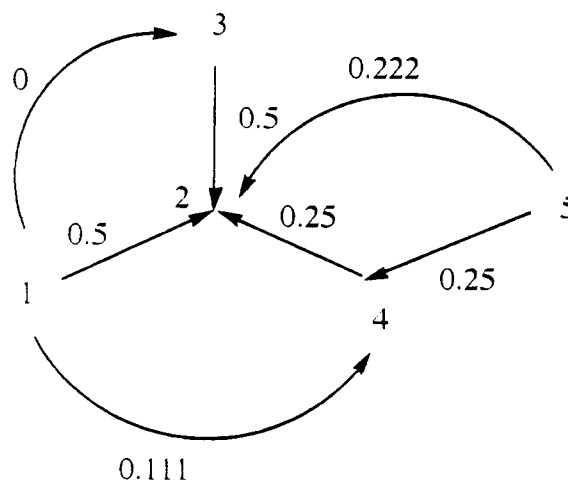
The existence of chemical information encoded in the different topological descriptors is evident. However, none of the more widely used indicators (Wiener, Balaban, Schultz, Randic, ...) describe in an explicit manner the charge distribution in the molecule, although Kier and Hall have developed an atomic value of chi capable of computing the number of valence electrons on the carbon atom.⁵

Since many important physical, chemical, and biological properties are related to the charge distribution, the introduction of a topological index able to characterize this essential property is necessary.

Since the charge influence decreases with the square of the distance, we use as a first charge descriptor the "inverse square distance matrix" (i.s.d.) which we denote by D^* . However, in order to avoid division by zero, the diagonal entries of the topological distance matrix remain the same.



A) Benzene



B) Isopentane

Figure 1. Charge term values for benzene and isopentane.

We now define the M matrix as

$$M = A \times D^*$$

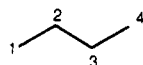
* Abstract published in *Advance ACS Abstracts*, March 1, 1994.

Table 1. Values of the G_k and J_k Charge Indexes up to the Fifth Order for a Set of Alkanes

molecule	N	G_1	G_2	G_3	G_4	G_5	J_1	J_2	J_3	J_4	J_5
3	3	0.5000	0.0000	0.0000	0.0000	0.0000	0.2500	0.0000	0.0000	0.0000	0.0000
4	4	0.5000	0.2222	0.0000	0.0000	0.0000	0.1667	0.0741	0.0000	0.0000	0.0000
2M3	4	1.5000	0.0000	0.0000	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000
5	5	0.5000	0.2222	0.1250	0.0000	0.0000	0.1250	0.0556	0.0313	0.0000	0.0000
2M4	5	1.5000	0.4444	0.0000	0.0000	0.0000	0.3750	0.1111	0.0000	0.0000	0.0000
22MM3	5	3.0000	0.0000	0.0000	0.0000	0.0000	0.7500	0.0000	0.0000	0.0000	0.0000
6	6	0.5000	0.2222	0.1250	0.0800	0.0000	0.1000	0.0444	0.0250	0.0160	0.0000
2M5	6	1.5000	0.4444	0.2500	0.0000	0.0000	0.3000	0.0889	0.0500	0.0000	0.0000
3M5	6	1.5000	0.6667	0.1250	0.0000	0.0000	0.3000	0.1333	0.0250	0.0000	0.0000
22MM4	6	3.0000	0.6667	0.0000	0.0000	0.0000	0.6000	0.1333	0.0000	0.0000	0.0000
23MM4	6	2.0000	0.8889	0.0000	0.0000	0.0000	0.4000	0.1778	0.0000	0.0000	0.0000
7	7	0.5000	0.2222	0.1250	0.0800	0.0556	0.0833	0.0370	0.0208	0.0133	0.0093
2M6	7	1.5000	0.4444	0.2500	0.1600	0.0000	0.2500	0.0741	0.0417	0.0267	0.0000
3M6	7	1.5000	0.6667	0.2500	0.0800	0.0000	0.2500	0.1111	0.0417	0.0133	0.0000
3E5	7	1.5000	0.6667	0.3750	0.0000	0.0000	0.2500	0.1111	0.0625	0.0000	0.0000
22MM5	7	3.0000	0.6667	0.3750	0.0000	0.0000	0.5000	0.1111	0.0625	0.0000	0.0000
23MM5	7	2.0000	1.1111	0.2500	0.0000	0.0000	0.3333	0.1852	0.0417	0.0000	0.0000
24MM5	7	2.5000	0.4444	0.5000	0.0000	0.0000	0.4167	0.0741	0.0833	0.0000	0.0000
33MM5	7	3.0000	1.1111	0.1250	0.0000	0.0000	0.5000	0.1852	0.0208	0.0000	0.0000
233MMM4	7	3.5000	1.3333	0.0000	0.0000	0.0000	0.5833	0.2222	0.0000	0.0000	0.0000
8	8	0.5000	0.2222	0.1250	0.0800	0.0556	0.0714	0.0317	0.0179	0.0114	0.0079
2M7	8	1.5000	0.4444	0.2500	0.1600	0.1111	0.2143	0.0635	0.0357	0.0229	0.0159
3M7	8	1.5000	0.6667	0.2500	0.1600	0.0556	0.2143	0.0952	0.0357	0.0229	0.0079
4M7	8	1.5000	0.6667	0.3750	0.0800	0.0556	0.2143	0.0952	0.0536	0.0114	0.0079
3E6	8	1.5000	0.6667	0.3750	0.1600	0.0000	0.2143	0.0952	0.0536	0.0229	0.0000
22MM6	8	3.0000	0.6667	0.3750	0.2400	0.0000	0.4286	0.0952	0.0536	0.0343	0.0000
23MM6	8	2.0000	1.1111	0.3750	0.1600	0.0000	0.2857	0.1587	0.0536	0.0229	0.0000
25MM6	8	2.5000	0.6667	0.2500	0.3200	0.0000	0.3571	0.0952	0.0357	0.0457	0.0000
2M3E5	8	2.0000	1.1111	0.6250	0.0000	0.0000	0.2857	0.1587	0.0893	0.0000	0.0000
223MMM5	8	3.5000	1.5556	0.3750	0.0000	0.0000	0.5000	0.2222	0.0536	0.0000	0.0000
234MMM5	8	2.5000	1.3333	0.5000	0.0000	0.0000	0.3571	0.1905	0.0714	0.0000	0.0000
9	9	0.5000	0.2222	0.1250	0.0800	0.0556	0.0625	0.0278	0.0156	0.0100	0.0069
4MS	9	1.5000	0.6667	0.3750	0.1600	0.0556	0.1875	0.0833	0.0469	0.0200	0.0069
235MMM6	9	3.0000	1.1111	0.5000	0.3200	0.0000	0.3750	0.1389	0.0625	0.0400	0.0000
24MM3E5	9	2.5000	1.3333	1.0000	0.0000	0.0000	0.3125	0.1667	0.1250	0.0000	0.0000
33EE5	9	3.0000	1.3333	0.7500	0.0000	0.0000	0.3750	0.1667	0.0938	0.0000	0.0000
2234M5	9	4.0000	1.7778	0.7500	0.0000	0.0000	0.5000	0.2222	0.0938	0.0000	0.0000
224MMM6	9	4.0000	0.8889	0.7500	0.2400	0.0000	0.5000	0.1111	0.0938	0.0300	0.0000
10	10	0.5000	0.2222	0.1250	0.0800	0.0556	0.0556	0.0247	0.0139	0.0089	0.0062

where A is the adjacency ($N \times N$) matrix. The charge terms, CT , are defined as $CT_{ij} = m_{ij} - m_{ji}$, where m stands for the elements of the M matrix.

To illustrate the calculation of the CT terms, let us consider the n -butane. We start by drawing the hydrogen-suppressed graph and then we number each vertex



Then:

$$A = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{vmatrix} \quad D^* = \begin{vmatrix} 0 & 1 & 1/4 & 1/9 \\ 1 & 0 & 1 & 1/4 \\ 1/4 & 1 & 0 & 1 \\ 1/9 & 1/4 & 1 & 0 \end{vmatrix}$$

Therefore

$$M = A \times D^* = \begin{vmatrix} 1 & 0 & 1 & 1/4 \\ 1/4 & 2 & 1/4 & 10/9 \\ 10/9 & 1/4 & 2 & 1/4 \\ 1/4 & 1 & 0 & 1 \end{vmatrix}$$

Thus

$$CT_{11} = CT_{44} = 1$$

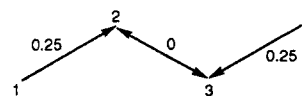
$$CT_{22} = CT_{33} = 2$$

$$CT_{14} = CT_{23} = 1/4 - 1/4 = 0$$

$$CT_{12} = -CT_{34} = 0 - 1/4 = -1/4$$

$$CT_{13} = -CT_{24} = 1 - 10/9 = -1/9$$

At this point, several facts should may be emphasized. For $i = j$ the CT_{ij} terms represent the topological valence of the vertex i (1,2,2,1 for the vertexes 1,2,3, and 4, respectively). For $i \neq j$ the CT_{ij} terms represent a measure of the net charge transferred from the atom "j" to "i". Thus, if the CT_{ij} is negative, the atom "i" will transfer the net charge to "j". Moreover, the CT_{ij} descriptors do not depend on the numerical order, i.e. they are really "graph invariants". The optimal topological illustration of these descriptors would be a directed graph (or digraph)



This implied charge distribution spectrum is surprisingly similar to the real charge distribution in the molecule.

Table 2. Linear Regression Equations Obtained To Correlate the Molecular Dipole Moment, μ (debyes), of a Set of Hydrocarbons with the Charge Indexes

variable no.	r	Cp	equation
1	0.677	62.30	$\mu = 0.200 \pm 2.535J_3$
2	0.727	53.00	$\mu = 0.116 + 0.203J_2 + 2.149J_3$
3	0.796	38.01	$\mu = 0.078 - 0.129G_2 + 0.569G_3 + 2.791J_2$
4	0.895	13.11	$\mu = 0.068 - 0.181G_2 + 1.389G_4 + 0.958J_2 + 0.103J_3$
5	0.922	6.98	$\mu = -0.053 - 0.139G_2 + 0.318J_1 + 0.858J_2 + 1.840J_3 + 7.212J_4$
6	0.941	3.53	$\mu = -0.037 - 0.179G_1 - 0.134G_2 + 1.365G_4 + 0.789J_1 + 0.820J_2 + 2.421J_3$
7	0.944	94.58	$\mu = -0.185 + 0.043N - 0.230G_1 - 0.163G_2 + 1.380G_4 + 0.917J_1 + 0.875J_2 + 2.392J_3$

Selected Equation

$\mu = -0.037 - 0.179G_1 - 0.134G_2 + 1.365G_4 + 0.789J_1 + 0.820J_2 + 2.421J_3$
 $n = 30 \quad r = 0.941 \quad SD = 0.107 \quad F = 28.0 \quad p < 0.001$

Table 3. Correlation of Molecular Dipole Moment Values, μ (debyes), of Hydrocarbons with Charge Indices^a

compound	observed ^b	calculated	residual
1,1-dimethylbutyne	0.66	0.71	0.05
propene	0.36	0.40	-0.03
acetylene	0.00	-0.04	0.04
ethylbenzene	0.59	0.60	-0.02
toluene	0.36	0.42	-0.06
propane	0.08	0.07	0.01
butane	0.04	0.04	0.00
2,2-dimethylpropane	0.00	0.02	-0.02
cyclopropene	0.45	0.39	0.06
propyne	0.78	0.73	0.05
propadiene	0.00	0.18	-0.18
cyclopropane	0.00	-0.04	0.04
cyclobutene	0.13	0.28	-0.15
1-butyne	0.80	0.76	0.04
1,2-butadiene	0.40	0.45	-0.05
1,3-butadiene	0.00	0.22	-0.22
1-butene	0.34	0.38	-0.04
2-methylpropene	0.50	0.41	0.09
2-methylpropane	0.13	0.09	0.04
1,3-cyclopentadiene	0.42	0.22	0.20
cyclopentene	0.20	0.15	0.05
1-pentene	0.81	0.82	-0.01
2-methyl-1,3-butadiene	0.25	0.33	-0.08
2-methylbutane	0.13	0.02	0.11
benzene	0.00	0.12	-0.12
1,3-cyclohexadiene	0.44	0.31	0.13
hexene	0.83	0.72	0.11
1,3,5-cycloheptatriene	0.25	0.15	0.10
o-xylene	0.62	0.64	-0.02

^a From selected equation in Table 2. ^b Data obtained from *Handbook of Tables of Organic Compounds Identification*, 52nd ed.; CRC Press: Cleveland, OH.

The calculation of the CT_{ij} indexes in the case of benzene allows for the characterization of aromaticity (see Figure 1A). In this case, as well as in all other unsaturated hydrocarbons, the multiple bonds have been counted as " n " edges in the graph as suggested, according to Randic and Murray.⁷

The corresponding values for n -hexane are

$$CT_{12} = -CT_{56} = -0.25$$

$$CT_{13} = -CT_{46} = -0.111$$

$$CT_{14} = -CT_{36} = -0.0625$$

$$CT_{15} = -CT_{26} = -0.04$$

$$CT_{16} = CT_{23} = CT_{24} = CT_{25} = CT_{34} = CT_{35} = CT_{45} = 0$$

It is evident, that the algebraic sum of the CT_{ij} terms, for $i > j$, will be null in the case of symmetrical compounds, which

implies (according to the ability of the CT_{ij} indexes to compute the charges distribution), that this sum would be proportional to the dipole moment of the molecule. In fact, all the symmetrical hydrocarbons, including those that are cyclic and aromatic, show dipole moments that are virtually null. However, nonsymmetrical molecules show low (non-null) dipole moments. For example, 2-methylbutane ($\mu = 0.13$ D) gives the values as shown in Figure 1.

In this case, i.e. for nonsymmetrical hydrocarbons, it is obvious that the algebraic sum of the CT_{ij} terms, for $i > j$, is not null. However, this sum would be dependent, in the case of branched molecules, on the numerical order of the different vertices; therefore, we may define the topological charge indexes, G_k , as

$$G_k = \sum_{i=1, j=i+1}^{i=N-1, j=N} |CT_{ij}| \delta(k, D_{ij})$$

where N is the number of vertices in the graph and D_{ij} values are the entries of the topological distance matrix. We may define k as the index order of G . Therefore, G_k would represent the sum of all the CT_{ij} terms, with $D_{ij} = k$. These new descriptors being "graph invariants" would evaluate the total charge transfer between atoms placed at topological distance " k ". Therefore, for a linear molecule there would exist $(N - 1)G_k$ values (from G_1 to G_{N-1}).

Furthermore, we may also introduce J_k as

$$J_k = G_k / (N - 1)$$

where N = number for atoms (or vertices). This index represents the mean value for the charge transfer for each bond (or edge), since the number of edges for alicyclic compounds is $N - 1$.

The G_k and J_k values, up to the fifth order for a set of alkanes, are listed in Table 1, together with the number of atoms, N .

The G_1 term is clearly related to the degree of branching, although many repeated values may be observed (0.5 for all the linear alkanes, 1.5 for alkanes containing tertiary carbons, and higher values for more branched hydrocarbons).

3. RESULTS

In order to check the ability of the G_k and J_k indexes to describe structures in such a way as to reveal a relationship with the molecular dipole moment, we selected a group of heterogeneous hydrocarbons to correlate this property. Table 2 illustrates the obtained multivariable linear regression equations for increasing the number of variables for a set of hydrocarbons containing mixed types of these.

Table 4. Linear Regression Equations Obtained To Correlate the Molecular Dipole Moment, μ (debyes), of a Set of Hydrocarbons with the Connectivity Indexes

variable no.	r	Cp	equation
1	0.315	0.45	$\mu = 0.057 + 0.083^0\chi$
2	0.510	-2.10	$\mu = -0.127 + 0.248^0\chi - 0.306^2\chi$
3	0.522	-0.43	$\mu = -0.145 + 0.280^0\chi - 0.390^2\chi$
4	0.522	1.42	$\mu = 0.106 + 0.268^0\chi - 0.399^2\chi + 0.495^4\chi_c + 0.165^4\chi_p$
5	0.534	3.20	$\mu = 0.243 + 0.423^0\chi - 0.389^1\chi - 0.790^3\chi_c - 0.209^4\chi_p + 1.712^4\chi_c$
Selected Equation			
$\mu = -0.243 + 0.423^0\chi - 0.389^1\chi - 0.790^3\chi_c - 0.209^4\chi_p + 1.712^4\chi_c$			
$n = 30 \quad r = 0.534 \quad SD = 0.262 \quad F = 1.83 \quad p < 0.146$			

Table 5. Linear Regression Equations Obtained To Correlate the Boiling Temperature, T_b ($^{\circ}\text{C}$), of a Set of Hydrocarbons with the Charge Indexes

variable no.	r	Cp	equation
1	0.987	129.9	$T_b = -119.2 + 29.2N$
2	0.994	36.0	$T_b = -103.5 + 28.6N - 36.1J_1$
3	0.995	27.0	$T_b = -112.0 + 29.5N - 8.1G_1 + 9.0G_2$
4	0.996	17.8	$T_b = -114.9 + 30.0N - 67.3G_1 - 30.5J_1 + 449.7J_3$
Selected Equation			
$T_b = -121.8 + 30.6N - 7.3G_2 - 74.0G_3 + 491.4J_3$			
$n = 39 \quad r = 0.997 \quad Cp = 5.0 \quad SD = 3.61 \quad F = 1299 \quad p < 0.001$			

As may be observed, Mallows lower Cp value is obtained for the six-variable equation, which leads us to choose this last equation as the most significant (with over 99.99% significance). The observed and predicted values are shown in Table 3. As a rule, good concordance is noted between observed and calculated values. Higher residuals were found for the null values; however their elimination produces nearly the same quality fit as the selected function. Although these results imply a significant achievement for the charge indexes in order with dipole moment prediction, its ability was compared to both the Wiener and connectivity indexes, χ_i , which may be considered as two of the most representative.

Table 4 illustrates the obtained correlations of dipole moment with the connectivity indexes for the same set of hydrocarbons. For a better criterion, a comparison of the statistics for the five-variable equations is shown.

$$\mu(D) = -0.053 - 0.139G_2 + 0.318J_1 + 0.858J_2 + 1.840J_3 + 7.212J_4$$

$$n = 30 \quad r = 0.922 \quad SD = 0.119 \\ F = 26.31 \quad p < 0.001$$

$$\mu(D) = -0.243 + 0.423^0\chi - 0.389^1\chi - 0.790^3\chi_c - 0.209^4\chi_p + 1.712^4\chi_c$$

$$n = 30 \quad r = 0.534 \quad SD = 0.262 \quad F = 1.83 \quad p < 0.146$$

While the correlation coefficient for the charge indexes is 0.922, its value is reduced to 0.534 with the χ_i terms, having virtually 100% probability in the first case and only 85% in the second. The standard error is reduced to more than 50%.

These results clearly demonstrate that the straightforward application of the charge indexes function has led to a considerable ability to predict the dipole moment value, which is all the more significant if we consider that the hydrocarbons are not polar molecules.

Table 6. Correlation of Boiling Temperature, T_b ($^{\circ}\text{C}$), of Hydrocarbons with Charge Indexes^a

molecule	observed ^b	calculated	residual
3	-42.07	-33.56	-8.51
4	-0.50	-0.61	0.11
2M3	-11.73	-10.23	-1.50
5	36.07	36.11	-0.04
2M4	27.85	25.02	2.83
22M3	9.5	9.45	0.05
6	68.74	63.70	5.04
2M5	60.27	61.73	-1.46
3M5	63.28	61.01	2.27
22MM4	49.74	47.02	2.72
23MM4	57.99	56.63	1.36
7	98.42	92.27	6.15
2M6	90.05	88.29	1.76
3M6	91.85	90.60	1.24
3E5	93.48	91.57	1.91
22MM5	79.20	80.61	-1.41
23MM5	89.78	91.56	-1.78
24MM5	80.50	82.77	-2.27
33MM5	86.03	83.24	2.79
223MMM4	80.88	80.92	-0.04
8	125.70	121.44	4.26
2M7	117.60	115.98	1.62
3M7	118.00	118.26	-0.30
4M7	117.70	117.84	-0.14
3E6	118.50	117.84	0.66
22MM6	106.80	106.87	-0.07
23MM6	115.60	118.79	-3.19
25MM6	109.00	110.99	-1.99
23ME5	115.60	117.68	-2.08
223MMM5	110.00	112.45	-2.45
234MMM5	113.40	116.25	-2.85
9	150.77	151.01	-0.24
4M8	142.40	145.18	-2.78
224MMM6	126.50	124.48	2.02
235MMM6	131.30	137.24	-5.94
24MM3E5	136.73	136.90	-0.17
33EE5	146.20	136.40	9.80
22MM7	133.00	133.71	-0.71
10	174.12	180.80	-6.68

^a From selected equation in Table 5. ^b Data obtained from *Handbook of Tables of Organic Compounds Identification*, 52nd ed.; CRC Press: Cleveland, OH.

A single charge index, J , was obtained as the sum of J_i , in order to develop a comparative study using the Wiener index, which also uses a single descriptor.

$$J = \sum_{i=1}^{i=5} J_i$$

The obtained equations are

$$\mu = -0.069 + 0.454J$$

$$n = 30 \quad r = 0.701 \quad SD = 0.203 \quad F = 26.1 \quad p < 0.01$$

and

$$\mu = 0.237 + 0.0045W$$

$$n = 30 \quad r = 0.338 \quad SD = 0.270 \quad F = 3.49 \quad p < 0.07$$

Against a correlation coefficient of 0.338 for W , a value of 0.701 is obtained for J , with 99% significance. This result is acceptable considering the structural heterogeneity of the selected group.

Since the charge distribution into the molecules, and particularly the dipole moment, closely determine the value and nature of the intermolecular forces, which may be reasonably probed by the boiling temperature and vaporization enthalpy, it was decided to correlate these two physicochemical properties with the charge indexes for a wide set of alkanes.

Table 7. Correlation of Boiling Temperature, T_b (°C), of Alcohols with Charge Indexes^a

molecule	observed ^a	calculated	residual
butanol	117.7	115.3	2.4
2-methylpropanol	107.9	110.0	-2.1
2-butanol	99.5	96.5	2.9
pentanol	137.8	135.9	1.9
2-methylbutanol	128.7	128.2	0.5
3-methylbutanol	131.2	129.6	1.6
2-pentanol	119.0	119.6	-0.6
3-pentanol	115.3	120.0	-4.6
3-methyl-2-butanol	111.5	111.9	-0.4
2-methyl-2-butanol	102.0	100.1	1.9
hexanol	157.0	155.7	1.3
2-hexanol	139.9	141.0	-1.1
3-hexanol	135.4	139.6	-4.2
3-methyl-3-pentanol	122.4	119.4	3.0
2-methyl-2-pentanol	121.4	126.6	-5.2
2-methyl-3-pentanol	126.5	132.7	-6.2
3-methyl-2-pentanol	134.2	131.8	2.4
3,3-dimethylbutanol	143.0	138.1	4.9
3,3-dimethyl-2-butanol	120.0	124.0	-4.0
4-methylpentanol	151.8	147.1	4.7
4-methyl-2-pentanol	131.7	134.7	-3.0
2-methylbutanol	146.5	148.8	-2.3
heptanol	176.3	175.2	1.1
2-methyl-2-hexanol	142.5	144.9	-2.4
3-methyl-3-hexanol	142.4	141.7	0.7
3-methyl-3-pentanol	142.5	140.5	2.0
2,3-dimethyl-2-pentanol	139.7	134.5	5.1
2,3-dimethyl-3-pentanol	139.0	133.3	5.7
2,4-dimethyl-2-pentanol	133.0	138.6	-5.6

^a Selected equation: $T_b = 47.85 + 18.43N - 6.34G_1 - 62.45J_2$; $n = 29$, $r = 0.979$, $SD = 3.63$, $F = 190.5$, $p < 0.001$. ^b Data obtained from *Handbook of Tables of Organic Compounds Identification*, 52nd ed.; CRC Press: Cleveland, OH.

The obtained results are not very significant, because, for nonpolar molecules the boiling point is dependent on the polarizability evaluated by London's dispersive forces and must therefore be dependent on the G_k . Thus, Table 5 illustrates the obtained results. Despite the low influence of the polar effects, a notable improvement is observed when charge indexes are included (compare with one and two variable equations).

As shown in Table 6 there is good concordance between observed and calculated values by using the five-variable regression equation.

However, if the study is applied to alcohols, calculating G_k by introducing the value 2.2 (which is Pauling's electronegativity of oxygen if chlorine is taken as pattern value = 2) in the oxygen entry of the main diagonal of the adjacency matrix, the results are much more significant.

$$T_b = 46.75 + 12.49N$$

$$n = 29 \quad r = 0.705 \quad SD = 12.16 \quad F = 26.09 \quad p < 0.01$$

and

$$T_b = 55.58 + 15.46N - 103.32J_2$$

$$n = 29 \quad r = 0.956 \quad SD = 5.09 \quad F = 138.8 \quad p < 0.001$$

In fact, the correlation coefficient increases from 0.705, using the correlation with N , to 0.956 using the two variables (N and J_2). This is in accordance with the polar character of alcohols, for which van der Waals forces play an important role. For the three-variable equation, the comparison between observed and calculated values is shown in Table 7.

For the vaporization enthalpy of alkanes, the addition of a second variable (G_1), clearly improves the regression quality

Table 8. Correlation of Vaporization Enthalpy, ΔH_v (kJ/mol), of Alkanes with Charge Indexes^a

molecule	observed ^b	calculated	residual
4	21.0	21.40	-0.40
2M3	19.30	19.96	-0.66
5	26.40	26.36	0.04
2M4	24.10	24.91	-0.81
22MM3	22.20	22.74	-0.54
6	31.60	31.30	0.30
2M5	29.80	29.86	-0.06
3M5	30.30	29.86	0.44
22MM4	27.70	27.69	0.01
23MM4	29.10	29.14	-0.05
7	36.60	36.25	0.35
2M6	34.90	34.81	0.09
3M6	35.10	34.81	0.29
3E5	35.50	34.81	0.69
22MM5	34.70	32.64	2.05
23MM5	32.50	34.09	-1.59
24MM5	35.10	33.37	1.73
33MM5	35.70	32.64	3.06
233MMM4	31.50	31.92	-0.42
8	41.60	41.20	0.40
2M7	39.60	39.76	-0.16
3M7	39.70	39.76	-0.06
22MM6	37.30	37.59	-0.29
23MM6	38.80	39.04	-0.24
25MM6	37.90	38.32	-0.42
2M3E5	38.60	39.04	-0.44
223MMM5	36.90	36.87	0.03
234MMM5	37.80	38.32	-0.62
9	46.50	46.15	0.35
33EE5	42.60	42.54	0.05
10	51.50	51.10	0.40
24MM6	37.80	38.32	-0.52
33MM6	37.50	37.59	-0.09
34MM6	39.00	39.04	-0.04
3M3E5	38.00	39.04	-1.04
224MM5	35.20	36.15	-0.95
233MMM5	37.20	36.87	0.33
2244MMMM5	38.50	38.94	-0.44
12	58.70	61.01	-1.31
16	81.30	80.81	0.49

^a Selected equation: $\Delta H_v = 2.318 + 4.951N - 1.444G_1$; $n = 40$, $r = 0.997$, $SD = 0.86$, $F = 2839$, $p < 0.0001$. ^b Data obtained from *Handbook of Tables of Organic Compounds Identification*, 52nd ed.; CRC Press: Cleveland, OH.

(from $r = 0.9836$, $SD = 1.91$ to $r = 0.9967$, $SD = 0.86$). The statistics of the two-variable equation, together with the set of observed and calculated values, are illustrated in Table 8.

4. DISCUSSION AND CONCLUSIONS

The introduction of these new "charge indexes", G_k and J_k , is of interest since it is a strictly topological quantity plausibly correlating with the charge distribution inside the molecule. This distribution is an important property which conditions the behavior of many physicochemical and biological properties.

The dipole moment has been chosen as the representative magnitude of the charge location, and its correlation with different topological indexes clearly demonstrates the descriptive ability of the G_k and J_k indexes compared with the χ_i connectivity or Wiener's path index, W . This result must be emphasized since the distance matrix may be obtained from the adjacency matrix and even though the three types of indexes are derived from this last matrix, nevertheless they are very different in encode chemical information.

The charge distributions descriptive ability has been also confirmed by the prediction of two properties closely related to it (by the intermolecular forces): boiling temperature and vaporization enthalpy.

As may be expected, in the case of alkanes, the new descriptors lead to significant but not spectacular improvements, with regard to the number of carbon atoms (N), due to the apolar character shown by the group. However, the advantage is evident when polar molecules (such as alcohols) are studied.

Although new tests must be developed in the future, the obtained results clearly point out that the charge indexes as new topological descriptors may be considered useful for describing the charge location inside the molecule.

ACKNOWLEDGMENT

The authors express their acknowledgment to Dr. José Jaen for his help on this paper redaction as well as to CICYT,

SAF92-0684 (Spanish Ministry of Science), for its economic support of our research development.

REFERENCES AND NOTES

- (1) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic: London, 1976.
- (2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
- (3) Randic, M.; Krais, G. A.; Deonauv-Jerman-Blazic, B. *Chemical Applications of Topology and Graph Theory*; Elsevier: Amsterdam, 1983; pp 192-205.
- (4) Klopman, G. *J. Am. Chem. Soc.*, **1984**, *106*, 7315.
- (5) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press, Ltd.: Letchworth, England, 1986; pp 69-75.
- (6) Mihalic, Z.; Nikolic, S.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28-37.
- (7) Murray, W. J.; Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1978**, *64*, 1978.