Journal of Mathematical Chemistry 9(1992)75–85

TOPOLOGICAL ATOMIC INDICES AND THE ELECTRONIC CHARGES IN ALKANES

Danail BONCHEV* and Lemont B. KIER
Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298, USA

Received 9 November 1990; revised 6 June 1991

Abstract

Various topological indices have been examined in the search for atomic descriptors reproducing the CNDO/2 electronic charges of the carbon atoms in alkanes. The best one-variable correlation ($r = 0.9902$, $s = 0.0034$ electron) was found with the number of nearest neighboring atoms. Two molecular connectivity indices introduced earlier by Hall and Kier have shown the highest two-parameter correlation ($r = 0.9992$, $s = 0.0009$ electron).

1. Introduction

The influence of molecular topology on the electronic charges of atoms can be studied best in hydrocarbons where all the skeleton atoms are the same chemical element and the structural patterns of the molecular skeletons are of prime importance. The progress in chemical graph theory [1–5] has revealed its applicability to practically all areas of chemistry, mainly by successful predictions of the properties and activities of molecules proceeding from their structure [6–10]. Useful atomic graph invariants have also been introduced, and some studies have related these invariants to NMR-chemical shifts and electronic charges [11–15].

Hall and Kier [11] reported in 1977 a high correlation between the CNDO/2 charges af alkanes and the molecular connectivity indices ${}^0\chi$, ${}^1\chi$, ${}^2\chi$ which are related to the simplest atomic graph-invariant, the number of the nearest neighboring atoms (or otherwise, the atom valence in the σ-skeleton or the vertex degree in terms of graph theory). Meanwhile, an important graph-invariant was found (the number of self-returning walks in molecular graphs) to be in a direct link with the method of moments in quantum mechanics [16–20]. This indicated a possibility to explain why topological indices work so well. Such an attempt has been reported [21] for one of the most successful topological indices, the Wiener index [22].

---

*On sabbatical leave from the Higher Institute of Chemical Technology, Burgas 8010, Bulgaria.

© J.C. Baltzer AG, Scientific Publishing Company

Related to the above-mentioned studies, the present work aims to reexamine the previous work of Hall and Kier [11] by comparing their results with those obtained by using different self-returning walks (SRW)-type indices and other related graph-invariants. The fact that the number of self-returning walks of length two for an atom equals the number of its nearest neighbors is a sufficient reason for such a study. On the other hand, this is a study of electron charges as computed with MO and is initiated with the study of CNDO/2 as one of the simpler theoretical models (and in view of the fact that earlier computations exist for comparison). The paper represents what may be called "structure–structure" correlation (in contrast to structure–property and property–property correlations). Here, a more convoluted property (CNDO/2 charge) is correlated with another theoretical property (topological indices) which has a more direct structural interpretation.

2. Atomic graph-invariants in molecules

Molecular connectivity is one of the most successful graph-theoretical concepts in chemistry. It quantifies specifically the branching and cyclicity of molecular carbon skeletons, as well as the degree of unsaturation and the influence of heteroatoms. The origin of the method is in the branching index proposed by Randić [23] and extended by Kier and Hall [6]. The hydrogen-suppressed graphs are used to represent the molecular skeleton. Each carbon atom is represented by a graph vertex. The *vertex degree* δ_i, which equals the number of atoms attached to atom i, is used in constructing basic graph-invariants. The three *atomic connectivity indices* used by Hall and Kier [11] are specified as follows:

$$^0\chi_i = \delta_i^{1/2}, \tag{1}$$

$$^1\chi_i = 1/2 \sum_s (\delta_i \delta_j)_s^{-1/2}, \tag{2}$$

$$^2\chi_i = 1/3 \sum_t (\delta_i \delta_j \delta_k)_t^{-1/2}, \tag{3}$$

where the summation in eqs. (2) and (3) is over all graph edges (bonds) s and over all two-edge paths t emanating from vertex i (fig. 1(a)).

The non-directed graph of 2-methylbutane shown in fig. 1(a) can also be drawn as a directed graph as shown in fig. 1(b). The latter illustrates the notion of *self-returning walk* of degree 2; this is a path which starts in a vertex, visits a neighboring vertex and turns back to the initial vertex. Clearly, the number of self-returning walks of length 2 for an atom equals the number of atoms attached to it: $SRW_i^2 = \delta_i$. For a molecule with N atoms as a whole, the total number of self-returning walks of length 2 equals the doubled number of edges (single bonds) in the graph, also termed the total adjacency A:

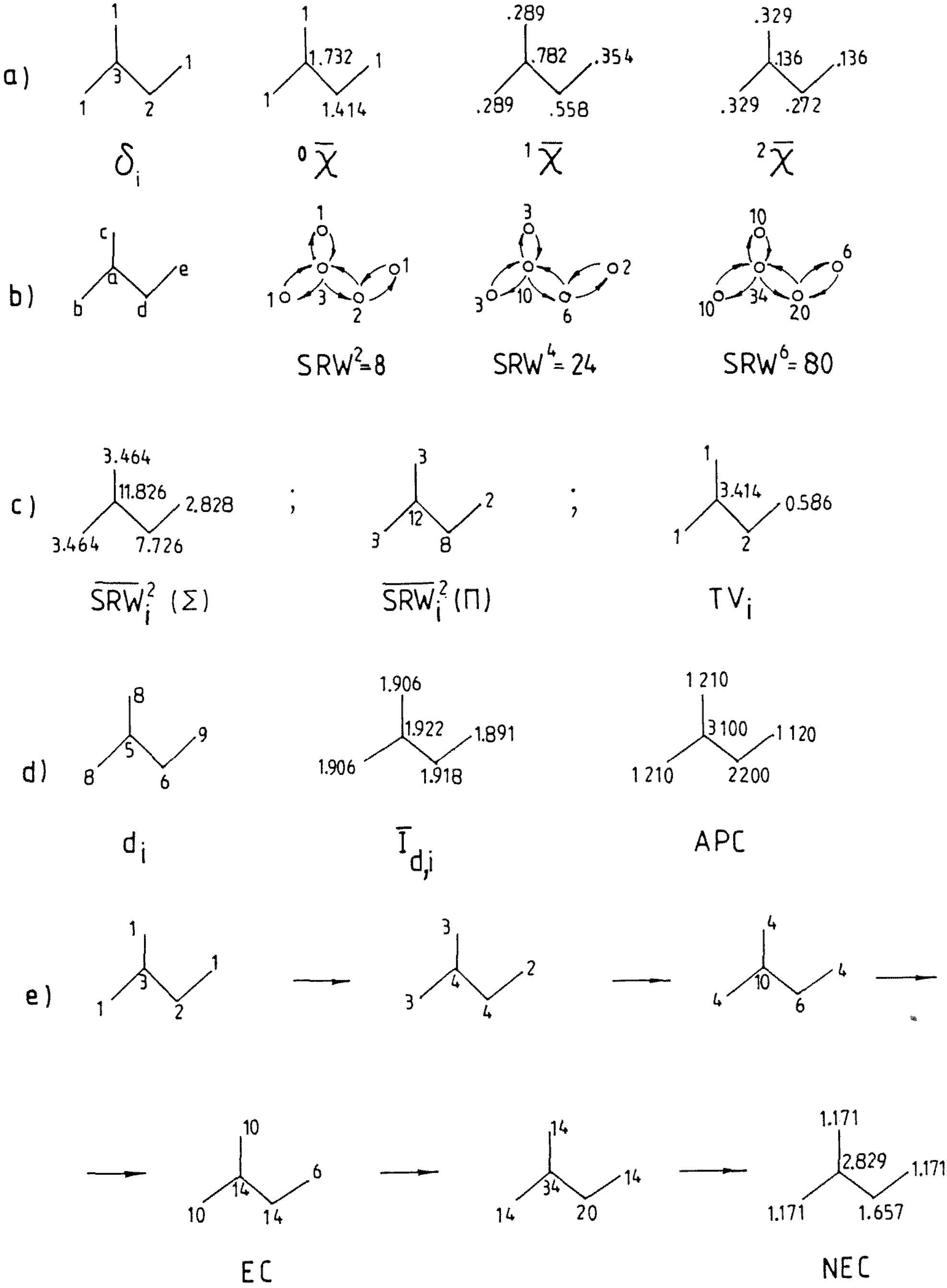


Fig. 1. (a) Zero-, first-, and second-order molecular connectivities as calculated from the vertex degree δ_i and eqs. (1) to (3). (b) Self-returning walks (*SRW*) of length 2, 4 and 6. (c) Two different weightings of *SRW*s (eq. (6)) and the topological charge (valency) (eq. (7)). (d) Distance number, information-theoretic index for the magnitude of graph distances, and Randić's atomic path code (eqs. (8) and (9)). (e) Extended connectivity (*RC*) and its normalized version (*NEC*).

$$SRW^2 = \sum_{i=1}^{N} SRW_i^2 = A. \tag{4}$$

Figure 1(b) also illustrates the number of self-returning walks of length 4 and 6 which can traverse repeatedly any of the edges they incorporate. SRW_i^4 and SRW_i^6 have also been examined as parameters for correlation with the CNDO/2 changes in our study, together with some derivative indices such as: $(SRW_i^4)^{1/2}$, $(SRW_i^4)^{1/4}$, $(SRW_i^6)^{1/3}$, $(SRW_i^6)^{1/6}$, SRW_i^6/δ_i^2, and SRW_i^6/δ_i^3. Each edge s included in these walks has been weighted by the geometric mean of the adjacent atoms $(\delta_i\, \delta_j)_s^{1/2}$. Being, however, traversed twice (in the forward and reverse direction), its total weight is either a sum or a product of the initial term. The total weight of all SRW_i^n of length n for the atom i is the sum over the weights of the individual $SRW_{i,k}^n$ (fig. 1(c)):

$$SRW_i^n = \sum_k SRW_{i,k}^n \tag{5}$$

and two different weightings were tried:

$$SRW_{i,k}^n = 2 \sum_{s \in k} (\delta_i\, \delta_j)_s^{1/2}; \quad SRW_{i,k}^n = \prod_{s \in k} (\delta_i\, \delta_j)_s\,. \tag{6}$$

The last atomic index, directly related to the self-returning walks, examined in this study was the so-called *topological charge, TC,* [24], specified by the authors in the following way:

$$TC = \lim_{n \to \infty} \frac{SRW_i^n}{SRW^n} \cdot SRW^2 = f_i N_\sigma\,. \tag{7}$$

Here, the number of self-returning walks of length 2 is substituted by the total number of σ-electrons taking part in the molecular σ-skeleton. The fractional topological charge f_i is a measure of the part of the time an electron moves near the atom i. As can be calculated from fig. 1(b), $f_c = f_b = 1/8$, $f_d = 1/4$, $f_a^{(2)} = 0.375$, $f_a^{(4)} = 0.417$, $f_a^{(6)} = 0.425, \ldots, f_a = 0.427$; $f_e^{(2)} = 0.125$, $f_e^{(4)} = 0.083$, $f_e^{(6)} = 0.075, \ldots, f_e = 0.073$. By multiplying by the number of σ-electrons ($N_\sigma = 8$), one arrives at the values shown in the last part of fig. 1(c).

Four more parameters have been examined as a result of a parallel study [24] which has shown that two topological indices, *EC* and *APC*, order the atoms in a molecule quite similarly to the ordering produced by SRW_i^n. On the other hand, the above-mentioned study revealed atom centricity to be the second major factor in the atom ordering besides its σ-skeleton valency δ_i. The best measure of the atom centricity (although a reverse proportional one) seems to be *the distance number of the vertex* d_i [25] which is the sum of the distances to all the remaining $N - 1$ atoms in the molecule. (The graph distance is an integer which is equal to the number of bonds along the shortest path between the two atoms):

$$d_i = \sum_{j=1}^{N-1} d_{ij} . \tag{8}$$

The *information-theoretic index* $I_{d,i}$, based on the distribution of the distances $d_{i,j}$ according to their magnitude [26,27] as presented by the Shannon-type formula:

$$I_{d,i} = -\sum_j d_{ij}/d_i \ \log_2 d_{ij}/d_i , \tag{9}$$

was also supposed to reflect the difference in the central to terminal location of atoms. The Randić atomic path code [28,29] *APC* is a four-digit code describing the number of the first, second, third, and fourth neighboring atoms. It is illustrated in fig. 1(d), together with d_i and $I_{d,i}$ values.

The next parameter checked was *EC*, the *extended connectivity*, of Morgan [30], which was found in the parallel study [24] to match the atom ordering produced by the number of self-returning walks. This was quite a surprise because *EC* has never been examined as a topological index. It was actually introduced to number uniquely the atoms within the coding procedure for chemical compounds, adopted initially by the Chemical Abstract Service. The extended connectivity of an atom is specified as the sum of the connectivities of the neighboring atoms in an iterative procedure which ends when the same atom ordering results in two consecutive iterations. This procedure has some deficiencies, discussed in the literature, namely, oscillations could occur in a pair of consecutive steps without arriving at a convergence. Such is the case with the molecular graph used to illustrate the topological indices under examination (fig. 1(e)). We actually modified the original scheme by introducing a *normalized extended connectivity, NEC*; to avoid difficulties with the different limits of convergence for the different atoms:

$$\delta_i = EC_i^{(n)} = \sum_{j \ \mathrm{adj} \ i} \delta_j^{(n-1)}, \tag{10}$$

$$NEC_i = \lim_{n \gg 1} \left(EC_i^{(n)} / \sum_i EC_i^{(n)} \right) \sum EC_i^{(1)}, \tag{11}$$

where $\sum_i EC_i^{(1)} = A = N_\sigma$ from eqs. (4) and (7), respectively.

The last topological index examined was the electrotopological state, E-state, recently introduced by Kier and Hall [31]. This is calculated from the intrinsic state value of atom i, I_i, plus the sum of loge state values $(I_j - I_i)/r^2$ vectored from atom i which account for the influence of molecular topology:

$$EST_i = I_i + \sum_j (I_j - I_i)/r_{ij}^2 . \tag{12}$$

I_j is the intrinsic state value for every other atom and r_{ij} is the graph distance within the loge in which i and j are terminal atoms. The intrinsic state value $I_i = (\delta_{v,i} + 1)/\delta_i$ is a count of valence electrons on a covalently bound atom plus 1, divided by the vertex degree of the atom.

3. Results

3.1. ONE-VARIABLE CORRELATIONS

All the parameters described in section 2 have been examined using a standard program for linear regression analysis with one, two, and three parameters. The first eleven alkane molecules from the reference publication [11] have been examined. They include all C_3 to C_6 alkanes with thirty-three nonequivalent carbon atoms (see table 2). The best ten parameters showing the highest correlation are collected in table 1. They are characterized by correlation coefficients $r > 0.95$. The two best indices–valency in the carbon skeleton (or the number of self-returning walks of length 2) and the Kier and Hall electrotopological state correlate with $r = 0.99$ and standard deviation of only 3.4×10^{-3} electrons, the atomic path code was third best with slightly worse statistics ($r = 0.986$, $s = 4.0$). All three parameters mentioned have also high Fisher ratio F and T-test. The remaining seven best indices are the two molecular connectivity indices of Hall and Kier [11], the weighted self-returning walks of length 2, two indices derivable from the number of self-returning walks of length 4, the topological charge, and the normalized extended connectivity. No correlation has been found for the information index on the distance distribution

Table 1

The ten atomic topological indices correlating linearly the best with the CNDO/2 charges, and their statistics

No.	Variables	r	$s^{a)}$	F	T
1	$\delta_i = SRW_i^2$	0.990	3.4	1552	27–39
2	EST	0.990	3.4	1551	39–43
3	APC	0.986	4.0	1073	23–32
4	${}^0\chi$	0.979	4.9	710	26–28
5	${}^1\chi$	0.973	5.5	556	17–23
6	$SRW_i^2\,(\Sigma)$	0.963	6.4	399	12–20
7	$(SRW_i^4)^{1/2}$	0.953	7.3	304	14–17
8	$(SRW_i^4)^{1/4}$	0.952	7.3	301	15–17
9	TC_i	0.952	7.3	301	10–13
10	NEC	0.952	7.4	298	12–17

a) In 10^{-3} electrons

($r = 0.003$), ${}^2\chi$ index ($r = 0.16$) and distance number ($r = 0.31$). The best three linear regressions are:

$$q_i = -35.53(\pm 1.31) - 25.49(\pm 0.65)\delta_i, \qquad (13)$$

$$q_i = 74.28(\pm 1.71) - 39.25(\pm 1.00)EST, \qquad (14)$$

$$q_i = -40.71(\pm 1.72) + 0.00261(\pm 0.0008)APC. \qquad (15)$$

The molecular connectivity index ${}^0\chi$ reported earlier as the best single variable is fourth best with the same statistics as in the reference paper [11], where it was determined for a sample of 193 atoms ($r = 0.979$, $s = 4.9$ versus 4.5×10^{-3} electrons).

3.2. TWO-VARIABLE CORRELATIONS

All pairs of variables have been examined, eliminating many of them which intercorrelated with $r > 0.8$. The best two-variable equations found are:

$$q_i = -65.75(\pm 0.80) + 128.16(\pm 1.00)\ {}^1\chi + 52.50(\pm 1.64)\ {}^2\chi$$
$$r = 0.9992,\ s = 0.94,\ F = 9963,\ T = 83,\ 128,\ 32; \qquad (16)$$

$$q_i = -37.61(\pm 1.23) + 24.80(\pm 0.57)\delta_i + 21.67(\pm 5.73)/d_i$$
$$r = 0.9933,\ s = 2.8,\ F = 1117,\ T = 31,\ 43,\ 4; \qquad (17)$$

$$q_i = -43.87 + 0.02516(\pm 0.00058)APC + 32.52(\pm 5.69)/d_i$$
$$r = 0.9933,\ s = 2.8,\ F = 1110,\ T = 33,\ 43,\ 6. \qquad (18)$$

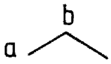
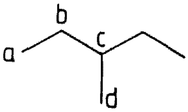
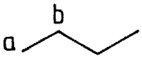
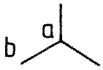
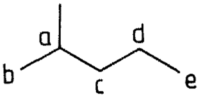
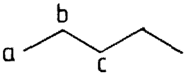
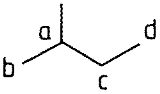
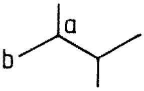
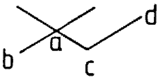
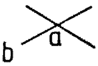
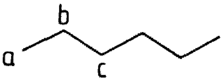
The intercorrelation between the three pairs of variables is: ${}^1\chi, {}^2\chi$ – 0.59; δ_i, d_i – 0.50 and APC, d_i – 0.48. The best correlation for two terms previously reported [11] dealt with ${}^0\chi$ and ${}^1\chi$. This pair of variables shows slightly worse statistics than those of eqs. (16) to (18) ($r = 0.990$, $s = 3.2$), but its intercorrelation is high (0.94).

The anticipated very favorable combination of the single variable, the σ-skeleton valency, δ_i, and the vertex distance number, was classified second. Considerably better statistics have been found for the pair of connectivity parameters ${}^1\chi$ and ${}^2\chi$ proposed by Hall, Kier, Murray and Randić (table 2). The correlation coefficient of 0.9992 and standard deviation of only 0.94×10^{-3} electrons, together with the very high F- and T-tests, make eq. (16) a reliable tool for calculating the CNDO/2 charges of alkanes. As is shown in table 3, the extrapolation made for higher alkanes also reproduces quite satisfactorily their charges.

The excellent statistics of eq. (16) makes the search for three-variable correlations unnecessary. In fact, a number of such correlations have been obtained, including a third parameter which does not intercorrelate with ${}^1\chi$ and ${}^2\chi$. These third variables

Table 2

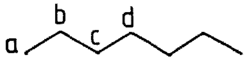
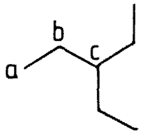
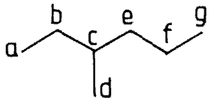
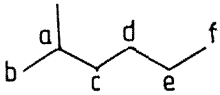
Comparison between CNDO/2 charges of alkanes and those calculated by a correlation with two topological indices

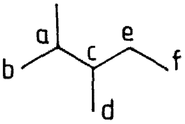
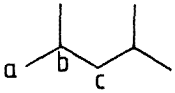
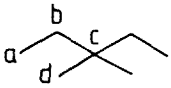
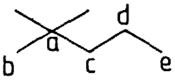
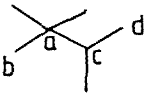
Alkane graph	Atom	CNDO/2	Eq. (16)	Alkane graph	Atom	CNDO/2	Eq. (16)
	a	−9	−8.0		a	−13	−13.2
	b	24	24.9		b	17	17.9
	a	−12	−11.6		c	39	37.9
	b	20	20.4		d	−12	−14.4
	a	46	45.2		a	40	39.5
	b	−9	−8.5		b	−11	−11.4
	a	−11	−11.6		c	17	15.5
	b	17	17.9		d	16	16.7
	c	17	15.8		e	−11	−11.6
	a	43	41.6		a	41	41.3
	b	−11	−11.4		b	−13	−12.8
	c	20	20.0		a	58	59.2
	d	−13	−13.2		b	−10	−10.0
	a	62	62.4		c	20	20.8
	b	−9	−7.5		d	−15	−14.2
	a	−11	−11.6				
	b	18	17.9				
	c	14	13.3				

proved to be the distance number d_i, the information index on the distance distribution $I_{d,i}$, the ratio SRW_i^6/σ_i^3, etc. They all have the same statistics: $r = 0.9993$, $s = 0.92 \times 10^{-3}$, $F = 6800$ to 7000, but the t-test indicates the marginal significance of the third parameter ($t = 1.3$ to 1.6). Our best two-variable correlation shows also

Table 3

Comparison between the CNDO/2 charges of some atoms in C_7 alkanes and those predicted by eq. (16)

		Carbon charge e/1000	
Alkane graph	Atom	CNDO/2	Eq. (16)
	a	−11	−11.7
	b	18	17.8
	c	14	13.3
	d	11	10.7
	a	−13	−13.3
	b	16	15.8
	c	37	34.2
	a	−13	−13.3
	b	17	17.9
	c	36	35.8
	d	−12	−14.5
	e	14	13.4
	f	16	16.7
	g	−11	−11.7
	a	40	39.5
	b	−11	−11.5
	c	14	12.9
	d	13	12.1
	e	18	17.8
	f	−11	−11.7

		Carbon charge e/1000	
Alkane graph	Atom	CNDO/2	Eq. (16)
	a	38	39.6
	b	−13	−12.8
	c	37	37.6
	d	−15	−15.8
	e	16	17.0
	f	−13	−13.3
	a	38	38.5
	b	−11	−11.5
	c	16	15.1
	a	−15	−14.3
	b	17	19.0
	c	54	56.0
	d	−12	−12.6
	a	55	57.4
	b	−11	−10.0
	c	16	16.3
	d	14	16.0
	e	−11	−11.7
	a	55	59.0
	b	−12	−11.2
	c	40	41.9
	d	−15	−13.6

better statistics than the best three-variable equation in the reference work [11] ($r = 0.9992$ versus 0.998, $s = 0.00094$ versus 0.00140), which includes the intercorrelated $^0\chi$ and $^1\chi$ indices.

The quality of eq. (16) is demonstrated in table 3. The calculated carbon electron charges of all 42 distinct carbon atoms of C_7 alkanes are shown to be reproduced fairly well by eq. (16), which was derived for C_3 to C_6 alkanes. The mean deviation of the predicted and CNDO/2 charges is 0.98×10^{-3}.

4. Concluding remarks

Reviewing the analyses of the best one-variable and two-variable regressions, one arrives at conclusions confirming the previous results of Hall and Kier. The major structural features that influence the CNDO/2 carbon electronic charges in alkanes are the type of carbon atoms (quaternary, tertiary, secondary and primary ones), as well as the type (and number) of their nearest neighbors – alpha and beta carbons. They are encoded in various ways in all the ten best variables in table 1. As can be judged from eq. (16), the first two of these factors which are encoded in the $^1\chi$ term are four times more significant than the influence of the beta-carbons which give the major contribution to the $^2\chi$ term. On the other hand, the centricity of the atom was also found to play an important role, as shown by eqs. (17) and (18).

Thus, the main value of the present study is not in finding a very accurate equation (eq. (16)) for fast calculation of the CNDO/2 charges of alkanes. Rather, it is in the methodological significance of the structural factors and respective graph invariants revealed to be important in charge calculations. Thus, the necessary basis is provided for calculations of electronic charges, otherwise specified by the more sophisticated ab initio methods [32].

Acknowledgements

Danail Bonchev wishes to express his appreciation to Dr. Lemont B. Kier for his hospitality, as well as to Jack W. Fraser for valuable discussions and for support of this work from Sterling Drug, Inc., Malvern, PA, and Allied Signal Corporation, Engineered Materials Research, Desplaines, IL.

References

[1] A.T. Balaban (ed.), *Chemical Application of Graph Theory* (Academic Press, London, 1976).
[2] N. Trinajstić, *Chemical Graph Theory*, Vol. 2, (CRC Press, Boca Raton, FL, 1983).
[3] A.T. Balaban, J. Chem. Inf. Comput. Sci. 25(1985)334.
[4] D.H. Rouvray, Sci. Amer. 254(1987)40.
[5] D.H. Rouvray (ed.), *Computational Chemical Graph Theory* (Nova Publ., New York, 1990).
[6] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[7] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis* (Research Studies Press, Chichester, UK, 1986).
[8] P.G. Seybold, M. May and U.A. Bagal, J. Chem. Educ. 64(1987)575.
[9] M.I. Stankevich, I.V. Stankevich and N.S. Zefirov, Usp. Khim. 57(1988)337.
[10] L.H. Hall, in: ref. [5], p. 201.
[11] L.H. Hall and L.B. Kier, Tetrahedron 33(1977)1953.

[12] M. Randić, Chem Phys. Lett. 58(1978)180.
[13] M. Randić, Int. J. Quant. Chem. 23(1983)1707.
[14] O. Mekenyan, A.T. Balaban and D. Bonchev, J. Magn. Res. 63(1985)1.
[15] M. Randić and N. Trinajstić, Theor. Chem. Acta 73(1988)233.
[16] F. Cyrrot-Lackmann, Ph.D. Thesis, Orsay (1968).
[17] J.K. Burdett, S. Lee and W.C. Sha, Croat. Chem. Acta 57(1984)1193.
[18] M. Randić, J. Comput. Chem. 1(1980)386.
[19] M. Randić, Siam J. Alg. Meth. 6(1985)145.
[20] J.K. Burdett, Acc. Chem. Res. 21(1988)189.
[21] D. Bonchev, O. Mekenyan and O.E. Polansky, in: *Graph Theory and Topology in Chemistry*, ed. R.B. King and D.H. Rouvray (Elsevier, Amsterdam, 1987), p. 126.
[22] H. Wiener, J. Amer. Chem. Soc. 69(1947)17.
[23] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.
[24] D.Bonchev, L.B. Kier and O. Mekenyan, Int. J. Quant. Chem., submitted.
[25] O.E. Polansky and D. Bonchev, MATCH 21(1987)135.
[26] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Chichester, UK, 1983).
[27] J. Klopman and R. Raychaudhury, J. Comput. Chem. 9(1988)232.
[28] M. Randić, J. Chem. Inf. Comput. Sci. 18(1978)101.
[29] M. Randić, MATCH 7(1979)65.
[30] H.L. Morgan, J. Chem. Doc. 5(1965)107.
[31] L.B. Kier and L.H. Hall, Pharm. Res., in press.
[32] D. Bonchev, L.H. Hall and L.B. Kier, work in progress.